



**BEST AVAILABLE COPY**

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In application of:	) Examiner: Spector, Lorraine
David BOTSTEIN, et al.	)
	) Art Unit: 1647
	)
Application Serial No. 09/989,721	) Confirmation No: 2434
	)
Filed: November 19, 2001	) Attorney's Docket No. 39780-2730 P1C55
	)
For: <b>NUCLEIC ACIDS ENCODING PRO809</b>	) <b>Customer No. 35489</b>

**EXPRESS MAIL LABEL NO. : EV 582 623 706 US**  
**DATE MAILED: NOVEMBER 22, 2005**

**ON APPEAL TO THE BOARD OF PATENT APPEALS AND INTERFERENCES**  
**APPELLANTS' BRIEF**

**MAIL STOP APPEAL BRIEF - PATENTS**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, Virginia 22313-1450

Dear Sir:

This Appeal Brief, filed in connection with the above captioned patent application, is responsive to the Final Office Action mailed on February 23, 2005. A Notice of Appeal was filed on July 22, 2005. A request for a two months extension of time is requested herewith. Appellants hereby appeal to the Board of Patent Appeals and Interferences from the final rejection in this case.

An Amendment After Final Rejection is being filed concurrently with the present appeal brief.

The following constitutes the Appellants' Brief on Appeal.

11/28/2005 CNGUYEN 00000029 081641 09989721  
01 FC:1252 450.00 DA  
11/28/2005 CNGUYEN 00000029 081641 09989721  
02 FC:1402 500.00 DA

### **I. REAL PARTY IN INTEREST**

The real party in interest is Genentech, Inc., South San Francisco, California, by an assignment of the parent application, U.S. Patent Application Serial No. 09/941,992 recorded November 16, 2001, at Reel 012176 and Frame 0450.

### **II. RELATED APPEALS AND INTERFERENCES**

The claims pending in the current application are directed to a polypeptide referred to herein as "PRO809." There exist two related patent applications, (1) U.S. Patent Application Serial No. 09/991,073, filed November 14, 2001 (containing claims directed to PRO809 polypeptides), and (2) U.S. Patent Application Serial No. 09/990,726, filed November 14, 2001 (containing claims directed to antibodies to the PRO809 polypeptide). These two related applications are also under final rejection from the same Examiner and based upon the same outstanding rejection, therefore appeal of these final rejections are being pursued independently and concurrently herewith.

### **III. STATUS OF CLAIMS**

Claims 122-126, 129-131 and 135-138 are in this application.

Claims 1-123, 125-128 and 132-134 have been canceled.

Claims 124, 129-131 and 135-138 stand rejected and Appellants appeal the rejection of these claims.

A copy of the rejected claims in the present Appeal is provided in the Claims Appendix.

### **IV. STATUS OF AMENDMENTS**

The claims involved in the appeal have been amended by an amendment filed concurrently with this appeal brief to cancel Claims 122-123 and 125-126. The claims listed in the Appendix incorporate this amendment.

### **V. SUMMARY OF CLAIMED SUBJECT MATTER**

The invention claimed in the present application is related to an isolated nucleic acid comprising: an isolated nucleic acid of SEQ ID NO: 222; the full-length coding sequence of the nucleic acid sequence of SEQ ID NO: 222; or the full-length coding sequence of the cDNA

-2-

On Appeal to the Board of Patent Appeals and Interferences

Appellants' Brief

Application Serial No. 09/989,721

Attorney's Docket No. 39780-2730 P1C55

deposited under ATCC accession number 203025 (independent Claim 124). The PRO809 gene was shown for the first time to be significantly amplified in human lung cancers as compared to normal, non-cancerous human tissue controls (Ct values of 1.05 -1.61, that is, **2.070 to 3.053** fold amplification). This is set forth in the specification, at least in the 'Gene Amplification assay,' Example 170, page 539, line 19, to page 555, line 5 (specifically, see Table 9A, page 550). The profiles of various primary lung tumors used for screening the PRO polypeptide compounds of the invention in the gene amplification assay are summarized on Table 8, page 546 of the specification. This feature is carried by all claims dependent directly or indirectly from Claim 124, namely, Claims 129-131 and 135-138. Methods for selecting a host are generally set forth in the specification at, for example, in Examples 140-143 and page 376, line 12 onwards (Claims 137-138), and describes the expression of PRO nucleic acids in various host cells, including *E. coli*, yeast and Baculovirus-infected insect cells. Methods for selecting a vector are generally set forth in the specification at, for example, on page 378, line 8 (Claims 135 and 136).

Finally, the presently claimed nucleic acid sequence (referred to in the present application as "DNA59841-1460") is shown in the present specification as SEQ ID NO: 222 while the polypeptide encoded by this polynucleotide sequence is defined as SEQ ID NO: 223, and is further described in Figures 150 and 151, respectively. The isolation of cDNA clones encoding PRO809 of SEQ ID NO: 222 is described at least in pages 140-142 and in Example 64, page 454 of the specification.

## **VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL**

1. Whether Claims 122-126, 129-131 and 135-138 satisfy the utility/ enablement requirement under 35 U.S.C. §§101/112, first paragraph.
2. Whether Claims 122-124 and 135-138 satisfy the written description requirement under 35 U.S.C. §112, first paragraph.
3. Whether Claims 122-126, 129-131 and 135-138 are patentable under 35 U.S.C. §102(b) over clone H74302 (EST isolated by Hillier *et al.*, 1995).

## VII. ARGUMENTS

### Summary of the Arguments

#### Issue 1: Utility/ Enablement

Claims 122-126, 129-131 and 135-138 stand rejected under 35 U.S.C. §§101/112, first paragraph, as allegedly lacking either a specific and substantial asserted utility or a well established utility.

The specification discloses that the gene encoding PRO809 showed amplification, ranging from **2.070 to 3.053** fold in different lung primary tumors. The Declaration of Dr. Audrey Goddard, submitted with Appellants' Response filed November 3, 2004, explains that a gene identified as being amplified at least 2-fold by the disclosed gene amplification assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy. Therefore, the gene amplification levels of **2.070 to 3.053** fold for lung primary tumors is considered significant. Hence, Appellants submit that one skilled in the art would find it more likely than not that PRO809 is useful as a diagnostic tool for detecting certain lung tumors.

Thus, Appellants submit that, as any skilled artisan in the field of oncology would easily appreciate, not all tumor markers are generally associated with every tumor, or even with most tumors. Therefore, whether the PRO809 gene is amplified in few tumor samples or in the vast majority of tumor samples studied is not relevant to its identification as a tumor marker, or its patentable utility. Rather, the fact that the amplification data for PRO809 is considered significant is what lends support to its usefulness as a tumor marker. Thus, a positive result does indicate the presence of cancer, while a negative result requires further follow up testing, testing which is considered routine by one skilled in the art of oncology and is not considered undue.

Appellants submit the art supports at least one utility for the PRO809 gene, that is, as a genetic biomarker for cancer or precancerous cells or damaged tissue. Appellants maintain that Sen *et al.* and the art support the Appellants' position that aneuploidy may be a feature of either cancerous or pre-cancerous tissue or damaged tissue. For example, Hittelman *et al.* shows that "epithelial tumors develop through a multistep process driven by genetic instability" in damaged lesions. Therefore, even if the observed PRO809 gene amplification were due to chromosomal



aneuploidy (which Appellants do not concede), it would still support at least one utility for the PRO809 gene according to Hittelman, because it helps in identifying individuals at significantly increased cancer risk. Appellants also submit that one skilled in the art would clearly know that early detection of lung cancer would provide information in advance about risk assessment, prognosis and therapy for lung cancer. Accordingly, the instant polynucleotides find utility as a diagnostic for cancer or for individuals at risk for developing lung cancer.

The Examiner further asserted on page 3 of the Final Office Action mailed February 23, 2004 that amplification of the PRO809 polynucleotide does not impart a specific, substantial, and credible utility since, “the literature cautions researchers from drawing conclusions based on small changes in transcript expression levels between normal and cancerous tissue.” In support of this assertion, the Examiner cited reference Hu *et al.*

First of all, the claims are directed to nucleic acids, not polypeptides, therefore, the issue of whether there is a correlation between gene amplification and polypeptide expression levels is irrelevant. One of skill in the art would understand how to use the claimed nucleic acids to detect amplification of the gene encoding PRO809, and how to use the gene amplification results to diagnose cancer. Thus, the question of whether or not PRO809 mRNA or polypeptide levels are also increased in these cancers has no relevance to the utility of the claimed nucleic acid molecules. Moreover, the teachings of Hu *et al.*, do not conclusively establish a *prima facie* case for lack of utility.

Instead, Appellants submit that based on the gene amplification data and the substantial, credible, asserted utility of the PRO809 gene in the diagnosis of lung cancer, one of ordinary skill would know exactly how to make and use these claimed nucleic acids for the diagnosis of cancers, without any undue experimentation.

## **Issue 2: Written Description**

Claims 122-124 and 135-138 stand rejected under 35 U.S.C. §112, first paragraph, as allegedly lacking adequate written description for the claimed variant polynucleotides.

Claims 122-123 have been canceled without prejudice or disclaimer. The polynucleotide sequence of SEQ ID NO: 222 claimed in Claim 124 has been reduced to practice.

Claims 135-138 depend from Claim 124 and therefore carry all the features of Claim 124. Methods for selecting a host are generally set forth in the specification, for example, in Examples 140-143 and page 376, line 12 onwards (Claims 137-138), which describes the expression of PRO nucleic acids in various host cells, including *E. coli*, yeast and Baculovirus-infected insect cells, and methods for selecting a vector are generally set forth in the specification at, for example, on page 378, line 8 (Claims 135 and 136). Accordingly, one skilled in the art would understand that the presently claimed invention of Claims 124 and 135-138 were reduced to practice at the time of filing of this application.

**Issue 3: Anticipation by Hillier *et al.*, 1995 (EST clone H74302)**

Claims 122-126, 129-131 and 135-138 are rejected under 35 U.S.C. §102(b) as being unpatentable over clone H74302 (EST isolated by Hillier *et al.*, 1995).

Appellants submit that the Hillier EST clone does not disclose the entire nucleotide sequence of SEQ ID NO: 222 nor did they disclose that their EST was part of cDNA (see enclosed alignment Item 1, Evidence Appendix). Thus, Hillier *et al.* did not teach nor reduce to practice the nucleotide sequence of SEQ ID NO: 222. Accordingly, one of skill in the art would not have been able to envision the cDNA clone defined in SEQ ID NO: 222 and therefore, Hillier *et al.* does not anticipate the instant invention.

These arguments are all discussed in further detail below under the appropriate headings.

**Response to Rejections**

**ISSUE 1. Claims 122-126, 129-131 and 135-138 are Supported by a Credible, Specific and Substantial Asserted Utility, and Thus Meet the Utility Requirement of 35 U.S.C. §§101/112, First Paragraph**

The sole basis for the Examiner's rejection of Claims 124, 129-131 and 135-138 under this section is that the data presented in Example 170 of the present specification is allegedly insufficient under the present legal standards to establish a patentable utility under 35 U.S.C. §101 for the presently claimed subject matter.

Claims 124, 129-131 and 135-138 stand further rejected under 35 U.S.C. §112, first paragraph, allegedly "since the claimed invention is not supported by either a specific and

substantial asserted utility or a well established utility for the reasons set forth above, one skilled in the art clearly would not know how to use the claimed invention."

Appellants strongly disagree and, therefore, respectfully traverse the rejection.

**A. The Legal Standard For Utility Under 35 U.S.C. § 101**

According to 35 U.S.C. § 101:

Whoever invents or discovers any new and *useful* process, machine, manufacture, or composition of matter, or any new and *useful* improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.  
(Emphasis added).

In interpreting the utility requirement, in *Brenner v. Manson*,<sup>1</sup> the Supreme Court held that the *quid pro quo* contemplated by the U.S. Constitution between the public interest and the interest of the inventors required that a patent applicant disclose a "substantial utility" for his or her invention, *i.e.*, a utility "where specific benefit exists in currently available form."<sup>2</sup> The Court concluded that "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion. A patent system must be related to the world of commerce rather than the realm of philosophy."<sup>3</sup>

Later, in *Nelson v. Bowler*,<sup>4</sup> the C.C.P.A. acknowledged that tests evidencing pharmacological activity of a compound may establish practical utility, even though they may not establish a specific therapeutic use. The Court held that "since it is crucial to provide researchers with an incentive to disclose pharmaceutical activities in as many compounds as possible, we conclude adequate proof of any such activity constitutes a showing of practical utility."<sup>5</sup>

---

<sup>1</sup> *Brenner v. Manson*, 383 U.S. 519, 148 U.S.P.Q. (BNA) 689 (1966).

<sup>2</sup> *Id.* at 534, 148 U.S.P.Q. (BNA) at 695.

<sup>3</sup> *Id.* at 536, 148 U.S.P.Q. (BNA) at 696.

<sup>4</sup> *Nelson v. Bowler*, 626 F.2d 853, 206 U.S.P.Q. (BNA) 881 (C.C.P.A. 1980).

<sup>5</sup> *Id.* at 856, 206 U.S.P.Q. (BNA) at 883.

In *Cross v. Iizuka*,<sup>6</sup> the C.A.F.C. reaffirmed *Nelson*, and added that *in vitro* results might be sufficient to support practical utility, explaining that "*in vitro* testing, in general, is relatively less complex, less time consuming, and less expensive than *in vivo* testing. Moreover, *in vitro* results with the particular pharmacological activity are generally predictive of *in vivo* test results, *i.e.*, there is a reasonable correlation there between."<sup>7</sup> The Court perceived, "No insurmountable difficulty" in finding that, under appropriate circumstances, "*in vitro* testing, may establish a practical utility."<sup>8</sup>

The case law has also clearly established that Appellants' statements of utility are usually sufficient, unless such statement of utility is unbelievable on its face.<sup>9</sup> The PTO has the initial burden to prove that Appellants' claims of usefulness are not believable on their face.<sup>10</sup> In general, an Applicant's assertion of utility creates a presumption of utility that will be sufficient to satisfy the utility requirement of 35 U.S.C. §101, "unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope."<sup>11, 12</sup>

---

<sup>6</sup> *Cross v. Iizuka*, 753 F.2d 1047, 224 U.S.P.Q. (BNA) 739 (Fed. Cir. 1985).

<sup>7</sup> *Id.* at 1050, 224 U.S.P.Q. (BNA) at 747.

<sup>8</sup> *Id.*

<sup>9</sup> *In re Gazave*, 379 F.2d 973, 154 U.S.P.Q. (BNA) 92 (C.C.P.A. 1967).

<sup>10</sup> *Ibid.*

<sup>11</sup> *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. (BNA) 288, 297 (C.C.P.A. 1974).

<sup>12</sup> See also *In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (C.C.P.A. 1980); *In re Irons*, 340 F.2d 974, 144 USPQ 351 (1965); *In re Sichert*, 566 F.2d 1154, 1159, 196 USPQ 209, 212-13 (C.C.P.A. 1977).

Compliance with 35 U.S.C. §101 is a question of fact.<sup>13</sup> The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the totality of the evidence under consideration.<sup>14</sup> Thus, to overcome the presumption of truth that an assertion of utility by the Appellant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Only after the Examiner made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the Appellant. The issue will then be decided on the totality of evidence.

The well established case law is clearly reflected in the Utility Examination Guidelines (“Utility Guidelines”),<sup>15</sup> which acknowledge that an invention complies with the utility requirement of 35 U.S.C. §101, if it has at least one asserted “specific, substantial, and credible utility” or a “well-established utility.” Under the Utility Guidelines, a utility is “specific” when it is particular to the subject matter claimed. For example, it is generally not enough to state that a nucleic acid is useful as a diagnostic without also identifying the conditions that are to be diagnosed.

In explaining the “substantial utility” standard, M.P.E.P. §2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase “immediate benefit to the public” or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be “currently available” to the public in order to satisfy the utility requirement. “Rather, any reasonable use that an Appellant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a “substantial” utility.”<sup>16</sup> Indeed, the Guidelines for Examination

---

<sup>13</sup> *Raytheon v. Roper*, 724 F.2d 951, 956, 220 U.S.P.Q. (BNA) 592, 596 (Fed. Cir. 1983) *cert. denied*, 469 US 835 (1984).

<sup>14</sup> *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d (BNA) 1443, 1444 (Fed. Cir. 1992).

<sup>15</sup> 66 Fed. Reg. 1092 (2001).

<sup>16</sup> M.P.E.P. §2107.01.

of Applications for Compliance With the Utility Requirement,<sup>17</sup> gives the following instruction to patent examiners: "If the Appellant has asserted that the claimed invention is useful for any particular practical purpose . . . and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility."

**B. Proper Application of the Legal Standard**

Appellants respectfully submit that the data presented in Example 170 starting on page 539 of the specification of the specification and the cumulative evidence of record, which underlies the current dispute, indeed support a "specific, substantial and credible" asserted utility for the presently claimed invention.

Example 170 describes the results obtained using a very well-known and routinely employed polymerase chain reaction (PCR)-based assay, the TaqMan<sup>TM</sup> PCR assay, also referred to herein as the gene amplification assay. This assay allows one to quantitatively measure the level of gene amplification in a given sample, say, a tumor extract, or a cell line. It was well known in the art at the time the invention was made that gene amplification is an essential mechanism for oncogene activation. Appellants isolated genomic DNA from a variety of primary cancers and cancer cell lines that are listed in Table 9 (pages 539 onwards of the specification), including primary lung and colon cancers of the type and stage indicated in Table 8 (page 546). The tumor samples were tested in triplicates with Taqman<sup>TM</sup> primers and with internal controls, beta-actin and GADPH in order to quantitatively compare DNA levels between samples (page 548, lines 33-34). As a negative control, DNA was isolated from the cells of ten normal healthy individuals, which was pooled and used as a control (page 539, lines 27-29) and also, no-template controls (page 548, lines 33-34). The results of TaqMan<sup>TM</sup> PCR are reported in  $\Delta$ Ct units, as explained in the passage on page 539, lines 37-39. One unit corresponds to one PCR cycle or approximately a 2-fold amplification, relative to control, two units correspond to 4-fold, 3 units to 8-fold amplification and so on. Using this PCR-based assay, Appellants showed that the gene encoding for PRO809 was amplified, that is, it showed

---

<sup>17</sup> M.P.E.P. §2107 II(B)(1).

approximately 1.05- 1.61  $\Delta$ Ct units for lung tumors which corresponds to  $2^{1.05}$ - $2^{1.61}$ - fold amplification in lung, or **2.070 to 3.053** fold in different lung primary tumors. Appellants submit that, one skilled in the art would find it more likely than not that PRO809 is useful as a diagnostic tool for detecting certain lung tumors.

However, the Examiner states regarding the teachings of the Goddard Declaration that “the issue....is *not* whether the technique is sensitive enough to detect a two-fold difference in amount of DNA, but rather that such was detected in only a minority of the tested lines of human lung tumor cell lines, which increase is likely to be due to aneuploidy in the tumorous tissue, and is neither diagnostic of cancer, not evidence of overexpression, which is the actual presence of extra protein encoded by the nucleic acid.” The Examiner further refers to Sen *et al.* (page 4 of the Final Office Action) to show that “cancerous tissues are known to be aneuploid” and that “pre-cancerous tissues *may* be aneuploid.” The Examiner further states that the Ashkenazi Declaration is not persuasive because in the instant case “it has not been established that the changes in the amount of DNA were significant. They occurred in only a minority of samples...and **cannot be considered to be diagnostic in the absence of significance, that is a statistically significant correlation between increased copy number and cancer.** It remains that random aneuploidy is the most parsimonious explanation of the results in the specification” (page 4, third paragraph, lines 18-24; emphasis added). Appellants respectfully disagree.

Appellants respectfully point out that the Declaration by Dr. Audrey Goddard presented in their response mailed provides a statement by an expert in the relevant art stating that “fold amplification” values of at least 2-fold are considered significant in the TaqMan™ PCR gene amplification assay. Appellants particularly draw the Board's attention to page 3 of the Goddard Declaration which clearly states that:

It is further my considered scientific opinion that an at least **2-fold increase** in gene copy number in a tumor tissue sample relative to a normal (*i.e.*, non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is **useful as a marker for the diagnosis of cancer**, for monitoring

cancer development and/or for measuring the efficacy of cancer therapy.  
(Emphasis added).

Accordingly, the **2.070 to 3.053** fold in the different lung primary tumors would be considered significant and credible by one skilled in the art, based upon the facts disclosed in the Goddard Declaration.

Further, as any skilled artisan in the field of oncology would easily appreciate, not all tumor markers are generally associated with every tumor, or even, with most tumors. In fact, some tumor markers are useful for identifying rare malignancies. That is, the association of the tumor marker with a particular type of tumor lesion may be rare, or, the occurrence of that particular kind of tumor lesion itself may be rare. In either event, even these rare tumor markers, which may not give a positive hit with most common tumors, have great value in tumor diagnosis, and consequently, in tumor prognosis. The skilled artisan would know that such tumor markers are very useful for better classification of tumors. Therefore, whether the PRO809 gene is amplified in a few lung tumors or in most tumors is not relevant to its identification as a tumor marker, or its patentable utility. Rather, whether the amplification data for PRO809 is significant is what lends support to its usefulness as a tumor marker. It was well known in the art at the time of filing of the application that gene amplification, which occurs in most solid tumors like lung and colon cancers, is generally associated with poor prognosis. Therefore, the PRO809 gene becomes an important diagnostic marker to identify such malignant lung or colon cancers, even if the malignancy associated with PRO809 molecule is a rare occurrence. Accordingly, the present specification clearly discloses enough evidence that the gene encoding the PRO809 polypeptide is significantly amplified in certain types of lung tumors and is therefore, a valuable diagnostic marker for identifying certain types of lung cancers.

Further, Appellants had submitted the Ashkenazi Declaration to show that “detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy.” Regarding Sen, Appellants agree that aneuploidy can be a feature of damaged tissue as well, besides cancerous or pre-cancerous tissue, and may not invariably lead to cancer. In fact Appellants maintain that Sen *et al.* support the Appellants’ position that aneuploidy may be a feature of either cancerous or pre-cancerous tissue or damaged



tissue. Further, the art shows that “epithelial tumors develop through a multistep process driven by genetic instability” in damaged lesions. Appellants provide a reference by Hittelman *et al.* (copy enclosed) to support this view. Hittelman studied damaged or premalignant lesions and suggests that epithelial tumors develop through a multistep process driven by genetic instability (see Hittelman abstract). Hittelman showed that a subset of the same molecular changes found in associated tumors were also found in premalignant lesions, suggesting that these premalignant lesions might represent precursor lesions for associated tumors, *i.e.*, a manifestation of a multistep tumorigenesis process. (See Hittelman, page 4, last three lines). Appellants submit that there is utility in identifying genetic biomarkers in epithelial tissues at cancer risk (also see Hittelman, abstract, line 4-7). Hittelman adds on page 2, fourth paragraph, line 3 that “it is important to identify individuals at significantly increased cancer risk who might best benefit from different types of intervention.”

Taken together, even if the observed PRO809 gene amplification were due to chromosomal aneuploidy (which Appellants do not concede), such an observation would still support at least one utility for the PRO809 gene according to Hittelman, because it helps in identifying individuals at significantly increased cancer risk. Therefore, the art supports at least one utility for the PRO809 gene, that is, as a genetic biomarker for cancer or precancerous cells or damaged tissue. As one skilled in the art would clearly know, early detection of lung cancer provides information in advance about risk assessment, prognosis and therapy for lung cancer. Accordingly, the instant polynucleotides find utility as a diagnostic for cancer or for individuals at risk for developing lung cancer.

Appellants also respectfully remind the Examiner that to overcome the presumption of truth that an assertion of utility by the Appellant enjoys, the Examiner must establish that it is “more likely than not” that one of ordinary skill in the art would doubt the truth of the statement of utility. The remarks in this rejection by the Examiner such as “ (instant nucleic acid) cannot be considered to be diagnostic in the absence of significance, that is a statistically significant correlation between increased copy number and cancer” (emphasis added) are a clear indication that the Examiner applies a standard that might be appropriate, if the issue at hand were the regulatory approval of a diagnostic assay based on the overexpression of PRO809 in lung tumor,

but is fully inappropriate for determining if the "utility" standard of the Patent Statute is met. The FDA reviewing an application for a new diagnostic assay will indeed ask for actual numerical data, statistical analysis, and other specific information before a diagnostic assay is approved. However, the Patent and Trademark Office is not the FDA, and the standards of patentability are not the same as the standards for market approval. It is well established law that therapeutic utility sufficient under the patent laws is not to be confused with the requirements of the FDA with regard to safety and efficacy of drugs to be marketed in the United States. *Scott v. Finney*, 34 F.3d 1058, 1063, 32 USPQ2d 1115, 1120 (Fed. Cir. 1994). Indeed, in *Nelson v. Bowler*, 626 F.2d 853, 856, 206 USPQ 881, 883 (CCPA 1980), the Federal Circuit found that the identification of a pharmacological activity of a compound provides an "immediate benefit to the public" and satisfies the utility requirement. This logically applies to a diagnostic utility as well. The identification of a diagnostic utility for a compound should suffice to establish an "immediate benefit to the public" and thus to establish patentable utility for PRO809.

**C. A prima facie case of lack of utility has not been established**

The Examiner further cited Hu *et al.*, to show that "the literature cautions researchers against drawing conclusions based on small changes in transcript expression levels between normal and cancerous tissues" (Page 3 of the Final Office Action mailed February 23, 2005).

First of all, as discussed above, the increase in DNA copy number for the PRO809 gene is significant according to the Goddard Declaration. Moreover, the instant invention is directed to polynucleotides, not to mRNA or polypeptides. Hence, the teachings of Hu are not relevant to the instant invention. Furthermore, even if it was relevant, contrary to the Examiner's assertion, the cited Hu *et al.* reference does not conclusively establish a *prima facie* case for lack of utility for the PRO809 molecule for the reasons outlined below.

The Hu *et al.* reference is entitled "Analysis of Genomic and Proteomic Data using Advanced Literature Mining" (emphasis added). Therefore, as the title itself suggests, the conclusions in this reference are based upon statistical analysis of information obtained from published literature, and not from experimental data. Hu *et al.* performed statistical analysis to provide evidence for a relationship between mRNA expression and biological function of a given

molecule (as in disease). The conclusions of Hu *et al.* however, only apply to a specific type of breast tumor (estrogen receptor (ER)-positive breast tumor) and cannot be generalized to breast cancer genes in general, let alone to cancer genes in general. Interestingly, the observed correlation was only found among ER-positive (breast) tumors not ER-negative tumors.” (See page 412, left column).

Moreover, the analytical methods utilized by Hu *et al.* have certain statistical drawbacks, as the authors themselves admit. For instance, according to Hu *et al.*, “*different statistical methods*” were applied to “*estimate the strength of gene-disease relationships and evaluated the results.*” (See page 406, left column, emphasis added). Using these different statistical methods, Hu *et al.* “[a]ssessed the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation.” (See page 411, left column). As is well known in the art, different statistical methods allow different variables to be manipulated to affect the resulting outcome. In this regard, the authors disclose that, “Initial attempts to search the literature” using the list of genes, gene names, gene symbols, and frequently used synonyms generated by the authors “revealed several sources of false positives and false negatives.” (See page 406, right column). The authors add that the false positives caused by “duplicative and unrelated meanings for the term” were “difficult to manage.” Therefore, in order to minimize such false positives, Hu *et al.* disclose that these terms “had to be eliminated entirely, thereby reducing the false positive rate but unavoidably under-representing some genes.” *Id.* (Emphasis added). Hence, Hu *et al.* had to manipulate certain aspects of the input data, in order to generate, in their opinion, meaningful results. Further, because the frequency of citation for a given molecule and its relationship to disease only reflects the current research interest of a molecule, and not the true biological function of the molecule, as the authors themselves acknowledge, the “[r]elationship established by frequency of co-citation do not necessarily represent a true biological link.” (See page 411, right column). Therefore, based on these findings, the authors add, “[t]his may reflect *a bias in the literature* to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently.” *Id.* (Emphasis added). In other words, some molecules may have been underrepresented merely because they were less frequently cited or

studied in literature compared to other more well-cited or studied genes. Therefore, Hu *et al.*'s conclusions are not based on genes/mRNA *in general*.

Appellants submit that, based on the nature of the statistical analysis performed herein, and in particular, based on Hu's analysis of only *one* class of genes, namely, the estrogen receptor (ER)-positive breast tumor genes, the conclusions drawn by the Examiner, namely that, "genes displaying a 5-fold change or less (mRNA expression) in tumors compared to normal showed no evidence of a correlation between altered gene expression and a known role in the disease (in general)" is not reliably supported.

In conclusion, when the proper legal standard is used, a *prima facie* case of lack of utility has not been met based on the cited reference Hu *et al.* by the Examiner.

Thus, based on the asserted utility for the PRO809 gene in the diagnosis of selected lung tumors, the reduction to practice of the instantly claimed nucleic acid sequence of SEQ ID NO:222 in the present application, the disclosure of the step-by-step protocols for making and isolating cDNA clones encoding PRO809 of SEQ ID NO:222 at least in Example 64, page 454 of the specification, the disclosure of a step-by-step protocol for expressing PRO809 cDNA in appropriate host cells (in Examples 140-143 and page 376, line 12), the step-by-step protocol of the gene amplification assay in Example 170, the skilled artisan would know exactly how to make and use the claimed nucleic acids for the diagnosis of lung cancers. That is, Appellants submit that based on the detailed information presented in the specification and the advanced state of the art in oncology, the skilled artisan would not have found any experimentation associated with testing lung tumors, given the present disclosure, 'undue.'

Therefore, since the instantly claimed invention is supported by either a credible, specific and substantial asserted utility or a well-established utility, and since the present specification clearly teaches one skilled in the art "how to make and use" the claimed invention without undue experimentation, Appellants respectfully request reconsideration and reversal of this outstanding rejections under 35 U.S.C. §101 and §112, first paragraph, to Claims 124, 129-131 and 135-138.

**ISSUE 2. Claims 122-124 and 135-138 Satisfy the Written Description Requirement of 35 U.S.C. §112, First Paragraph**

Claims 122-123 have been canceled without prejudice or disclaimer. The polynucleotide sequence of SEQ ID NO:222 claimed in Claim 124 has been reduced to practice and hence, this rejection directed to Claim 124 is improper. Claims 135-138 depend from Claim 124 and therefore carry all the features of Claim 124. Moreover, methods for selecting a host are generally set forth in the specification, for example, in Examples 140-143 and page 376, line 12 onwards (Claims 137-138), which describes the expression of PRO nucleic acids in various host cells, including *E. coli*, yeast and baculovirus-infected insect cells, and methods for selecting a vector are generally set forth in the specification at, for example, on page 378, line 8 (Claims 135 and 136). Accordingly, one skilled in the art would understand that the Appellants were in possession of the presently claimed invention of Claims 124 and 135-138 at the time of filing of this application.

Accordingly, Appellants request reconsideration and reversal of this outstanding rejections under 35 U.S.C. §112, first paragraph, for lack of written description.

**ISSUE 3. Claims 122-126, 129-131 and 135-138 are not Anticipated Under 35 U.S.C. §102(b) by Hillier *et al.*, 1995 (EST clone H74302)**

Claims 122-126, 129-131 and 135-138 are rejected under 35 U.S.C. §102(b) as being unpatentable over clone H74302 (EST isolated by Hillier *et al.*, 1995).

Appellants submit that the Hiller EST clone does not disclose the entire nucleotide sequence of SEQ ID NO: 222 nor did they disclose that their EST was part of a cDNA sequence (see enclosed alignment Item 1, Evidence Appendix). Thus, Hillier *et al.* did not teach nor reduce to practice the nucleotide sequence of SEQ ID NO:222 as disclosed in the present application. Further, one of skill in the art would not have been able to envision the cDNA clone defined in SEQ ID NO:222 from the teachings of Hillier and therefore, Hillier *et al.* does not anticipate the instant invention.

Accordingly, this rejection under 35 U.S.C. §102(b) should be withdrawn.

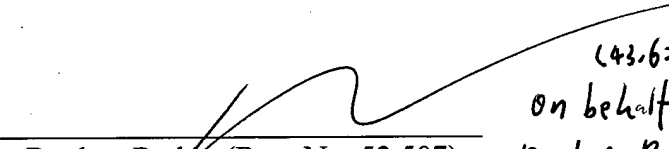
### CONCLUSION

For the reasons given above, Appellants submit that present specification clearly describes, details and provides a patentable utility for the claimed invention. Moreover, it is respectfully submitted that based upon this disclosed patentable utility, the present specification clearly teaches "how to use" the presently claimed nucleic acid. As such, Appellants respectfully request reconsideration and reversal of the outstanding rejection of Claims 122-126, 129-131 and 135-138.

The Commissioner is authorized to charge any fees which may be required, including extension fees, or credit any overpayment to Deposit Account No. 08-1641 (referencing Attorney's Docket No. 39780-2730 P1C55).

Respectfully submitted,

Date: November 22, 2005

By:   
Daphne Reddy (Reg. No. 53,507)

(43.626)  
on behalf of  
Daphne Reddy

**HELLER EHRMAN LLP**  
275 Middlefield Road  
Menlo Park, California 94025-3506  
Telephone: (650) 324-7000  
Facsimile: (650) 324-0638

## **VIII. CLAIMS APPENDIX**

### **Claims on Appeal**

124. An isolated nucleic acid comprising:

- (a) the nucleic acid sequence of SEQ ID NO: 222;
- (b) the full-length coding sequence of the nucleic acid sequence of SEQ ID NO: 222;

or

(c) the full-length coding sequence of the cDNA deposited under ATCC accession number 203025.

129. The isolated nucleic acid of Claim 124 comprising the nucleic acid sequence of SEQ ID NO: 222.

130. The isolated nucleic acid of Claim 124 comprising the full-length coding sequence of the nucleic acid sequence of SEQ ID NO: 222.

131. The isolated nucleic acid of Claim 124 comprising the full-length coding sequence of the cDNA deposited under ATCC accession number 203025.

135. A vector comprising the nucleic acid of Claim 124.

136. The vector of Claim 135, wherein said nucleic acid is operably linked to control sequences recognized by a host cell transformed with the vector.

137. A host cell comprising the vector of Claim 135.

138. The host cell of Claim 137, wherein said cell is a CHO cell, an *E. coli* or a yeast cell.

## **IX. EVIDENCE APPENDIX**

1. Alignment of the instantly claimed DNA 57836-1338 with EST sequence of Hillier *et al.*, (H74302).
2. Hittelman *et al.* 2001, Ann. N.Y. Acad. Sci. 952: 1-12.
3. Declaration of Audrey Goddard, Ph.D. under 35 C.F.R. §1.132, with attached Exhibits A-G:
  - A. Curriculum Vitae of Audrey D. Goddard, Ph.D.
  - B. Higuchi, R. et al., "Simultaneous amplification and detection of specific DNA sequences," *Biotechnology* 10:413-417 (1992).
  - C. Livak, K.J., et al., "Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization," *PCR Methods Appl.* 4:357-362 (1995).
  - D. Heid, C.A. et al., "Real time quantitative PCR," *Genome Res.* 6:986-994 (1996).
  - E. Pennica, D. et al., "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors," *Proc. Natl. Acad. Sci. USA* 95:14717-14722 (1998).
  - F. Pitti, R.M. et al., "Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer," *Nature* 396:699-703 (1998).
  - G. Bieche, I. et al., "Novel approach to quantitative polymerase chain reaction using real-time detection: Application to the detection of gene amplification in breast cancer," *Int. J. Cancer* 78:661-666 (1998).
4. Declaration of Avi Ashkenazi, Ph.D. under 35 C.F.R. §1.132.
5. Hu *et al.*, "Analysis of genomic and proteomic data using advanced literature mining," *Journal of Proteome Research* 2:405-412 (2003).
6. Sen S., "Aneuploidy and Cancer", *Current Opinion in Oncology*, 12: 82-88, (2000).

Item 1 is submitted herewith in response to the Examiner's request for an alignment of the claimed nucleic acids with the sequence of the EST clone (see page 6 of Final Office Action mailed February 23, 2005).

Item 2 is submitted herewith to support Appellants' assertion based on the art.

Items 3-4 were submitted with the Appellants' Response mailed on November 3, 2004.



Items 5-6 were made of record by the Examiner in the Office Action mailed June 30, 2005.

**X. RELATED PROCEEDINGS APPENDIX**

None.

SV 2147739 v1  
11/22/05 11:36 AM (39780.2730)

<first sequence: ss.H74302 (length = 582)  
<second sequence: ss.DNA57836 (length = 992)

<503 matches in an overlap of 562: 89.50 percent similarity  
<gaps in first sequence: 0, gaps in second sequence: 13 (20 bases)  
<score: 1385 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)  
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:36:56 2005]: 1 sequence found

H74302        yu56c11.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone  
IMAGE:230132 5', mRNA sequence. 582 bp,  
mRNA, linear, EST 31-OCT-1995

ACCESSION    H74302  
VERSION      H74302.1 GI:1047713  
KEYWORDS     EST; 5\_prime.  
SOURCE       Homo sapiens (human)

ORGANISM     Homo sapiens  
REFERENCE    1 (bases 1 to 582)

AUTHORS      Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,  
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.  
and Marra,M.

TITLE        Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL      Genome Res. 6 (9), 807-828 (1996)  
PUBMED       8889549

COMMENT      Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1114

High quality sequence stops: 308  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1114    Std Error: 0.00  
Seq primer: M13RP1

High quality sequence stop: 308.

FEATURES     Location/Qualifiers  
    source     1..582  
                /organism="Homo sapiens"  
                /mol\_type="mRNA"  
                /db\_xref="GDB:3781228"  
                /db\_xref="taxon:9606"  
                /clone="IMAGE:230132"  
                /sex="male"  
                /dev\_stage="20 week-post conception fetus"  
                /lab\_host="DH10B (ampicillin resistant)"  
                /clone\_lib="Soares fetal liver spleen 1NFLS"  
                /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
                with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
                1st strand cDNA was primed with a Pac I - oligo(dt) primer  
                [5' AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'],  
                double-stranded cDNA was ligated to Eco RI adaptors  
                (Pharmacia), digested with Pac I and cloned into the Pac I  
                and Eco RI sites of the modified pT7T3 vector. Library  
                went through one round of normalization. Library  
                constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

```

          10          20          30          40          50
ss.H74302      CAGGAAGTAGGAGGTTCTCACTGCCCCGAGCAGAXGGCCCTACACCCACCGA
                *****
ss.DNA57836    GGCACGAGCCAGGAAGTCTCACTGCCCCGAGCAGAGGGCCCTACACCCACCGA
                10          20          30          40          50

          60          70          80          90          100          110
ss.H74302      GGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTTCTCAA
                *****
ss.DNA57836    GGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGC-TTCTCAA
                60          70          80          90          100          110

          120          130          140          150          160          170
ss.H74302      GGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTT
                *****
ss.DNA57836    GGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTT
                120          130          140          150          160          170

          180          190          200          210          220          230
ss.H74302      CCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCAC
                *****
ss.DNA57836    CCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCAC
                180          190          200          210          220          230

          240          250          260          270          280          290
ss.H74302      CTATTCCTCTGTGGAACCAAGAACATCAAGGTTGCCAAGAAGGTGGTGAAGACCCACGA
                *****
ss.DNA57836    CTATTCCTCTGTGGAACCAAGAACATCAAGGTTGCCAAGAAGGTGGTGAAGACCCACGA
                240          250          260          270          280          290

          300          310          320          330          340          350
ss.H74302      GCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGGCTCACCTAAT
                *****
ss.DNA57836    GCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCT-GCTCACCTACT
                300          310          320          330          340          350

          360          370          380          390          400          410
ss.H74302      TTATGGCCGGGCGTCCTCCACCTXAGGTGCCCATGTGGGACAGTTXCCAGGGTTACAGAT
                * *****
ss.DNA57836    T--CTGCCGGGCGTCCTCCACCTCAGGTGCCCATGT-GGACAGT--GCCAGGCTACAGAT
                360          370          380          390          400          410

          420          430          440          450          460          470
ss.H74302      GCATTGGGGAGGTTGTGTTCCAAGCCAGTGTXTGAGGTTGCGGGGXCAATTTXAATTTT
                *** *
ss.DNA57836    GCACTGGG--AGCTGTGG-TCCAAGCCAGTGCTGAGCT--GCGGGCCAACCTTCACTCT
                420          430          440          450          460

          480          490          500          510          520          530
ss.H74302      GXAGGACAGAGGGGXAAAGGXCCAGGGTTXGGAGXTGATTTTGCCAGGXTTTTtagGGXAG
                * *****
ss.DNA57836    GCAGGACAGAGGGGCAGGCCCCAGGGT--GGAGATGA-TCTGCCAGGCGTCCTCGGGCAG
                470          480          490          500          510          520

          540          550          560          570          580
ss.H74302      TCCAATTATTXATCAAAAAGTTXATTTXGGGAAGXTTTGGAAGGTTXAAT
                *** *
ss.DNA57836    CCCACCTA-TCACCAACAGCCTGA--TCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAG
                530          540          550          560          570          580

          590          600          610          620          630          640
ss.DNA57836    ACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTT
                590          600          610          620          630          640

```

ss.DNA57836 CTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCC  
650 660 670 680 690 700

ss.DNA57836 AGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTT  
710 720 730 740 750 760

ss.DNA57836 GCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTTCAGGATAGG  
770 780 790 800 810 820

ss.DNA57836 GAATGGGGAGGTTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCC  
830 840 850 860 870 880

ss.DNA57836 AAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTTCGTATTTGGAGTTCATGCAA  
890 900 910 920 930 940

ss.DNA57836 AATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAA  
950 960 970 980 990

<first sequence: ss.H74303 (length = 414)  
<second sequence: ss.DNA57836 (length = 992)

<196 matches in an overlap of 414: 47.34 percent similarity  
<gaps in first sequence: 8 (20 bases), gaps in second sequence: 3 (6 bases)  
<score: 474 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)  
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:35:36 2005]: 1 sequence found

H74303      yu56c11.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone  
IMAGE:230132 3', mRNA sequence. 414 bp,  
mRNA, linear, EST 31-OCT-1995

ACCESSION    H74303  
VERSION      H74303.1 GI:1047714  
KEYWORDS     EST; 3\_prime.  
SOURCE       Homo sapiens (human)  
ORGANISM     Homo sapiens  
REFERENCE    1 (bases 1 to 414)  
AUTHORS      Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,  
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.  
and Marra,M.

TITLE        Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL      Genome Res. 6 (9), 807-828 (1996)  
PUBMED       8889549  
COMMENT      Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1114  
High quality sequence stops: 313  
Source: IMAGE Consortium; LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1114    Std Error: 0.00  
Seq primer: Promega -21m13  
High quality sequence stop: 313.

FEATURES  
source        Location/Qualifiers  
1..414  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:3781228"  
/db\_xref="taxon:9606"  
/clone="IMAGE:230132"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares fetal liver spleen 1NFLS"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

ss.DNA57836 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCAGCAGAGGCCCTACACCCACCGAG  
10 20 30 40 50 60

ss.DNA57836 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG  
70 80 90 100 110 120

ss.H74303 GTG

ss.DNA57836 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC  
130 140 150 160 170 180

ss.H74303 GCAAGAGCAGCTAAAACACACTCATTTGTCATGAACTCCAAATACGAACAGTGCACGCTG  
10 20 30 40 50 60  
\*\*\* \*\* \*\* \* \* \* \* \* \* \* \* \* \* \* \*

ss.DNA57836 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT  
190 200 210 220 230 240

ss.H74303 ATGGC--CTGCAGTCCTCTGCCGTGCTTGGCTCTCTGGACGGTTTCATTCTACATGGCTGC  
70 80 90 100 110 120  
\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

ss.DNA57836 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC  
250 260 270 280 290 300

ss.H74303 TGCTTTGCGTC-CTCTGACCTCCCCATTC---CCTATCCTGAACCCCCAAACTCCTCT  
130 140 150 160 170  
\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

ss.DNA57836 CGGCCTCCTTCAACCTCAACGTCACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT  
310 320 330 340 350 360

ss.H74303 TC-----ACTCAGACGGCGGGTGTCTCTGTAGAGCG-GCAAGGCAAGGATGGGGCT--C  
180 190 200 210 220  
\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

ss.DNA57836 GCCGGGCGTCTCCACCTCAGGTGCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG  
370 380 390 400 410 420

ss.H74303 TCCAGGGGACCCTGCCAGTCTCCATCTTCTGGTCACCACCTGGGGGACCACCTGTGAG-  
230 240 250 260 270 280  
\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

ss.DNA57836 AGCTGTGGTCCAAGCCAGTGTCTGAGTGCAGGCAACTTCACTCTGCAGGACAGAGGGG  
430 440 450 460 470 480

ss.H74303 --GGCGCTGTGCTGGACATTGGCGTTGTTTGCAGCCTGGCACCAGAACAGTCCGATGTC  
290 300 310 320 330 340  
\*\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

ss.DNA57836 CAGGCCCCAGGGTGGAGA-TGATCTGCCAGGCGTCTCGGGCAGCCACCTATCACCAAC  
490 500 510 520 530

ss.H74303 TGGCTCGGCAGGAAGGAGAAGTTGGCAGG-CTGCCTGTGGATXGGTCTCTGCTGCAGGTG  
350 360 370 380 390 400  
\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*

ss.DNA57836 AGCCTGATCGGGAAGGA----TGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCA  
540 550 560 570 580 590

ss.H74303 GACCTGCCCAT  
410  
\* \* \* \* \* \*

ss.DNA57836 G-CCTGCCAACTTCTCCTTCTGCCAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTG  
600 610 620 630 640 650

ss.DNA57836 CAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCAGGTGGTGACCAGA  
660 670 680 690 700 710

ss.DNA57836 AGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGA  
720 730 740 750 760 770

ss.DNA57836 GCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTTCAGGATAGGGAATGGGGAGGTCA  
780 790 800 810 820 830

ss.DNA57836 GAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGG  
840 850 860 870 880 890

ss.DNA57836 ACTGCAGGCCATCAGCGTGCACTGTTTCGTATTTGGAGTTCATGCAAATGAGTGTGTTTT  
900 910 920 930 940 950

ss.DNA57836 AGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAA  
960 970 980 990



<first sequence: ss.H73373 (length = 480)  
<second sequence: ss.DNA57836 (length = 992)

<456 matches in an overlap of 471: 96.82 percent similarity  
<gaps in first sequence: 0, gaps in second sequence: 7 (9 bases)  
<score: 1303 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)  
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:35:00 2005]: 1 sequence found

H73373        yu48f10.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone  
IMAGE:229387 5', mRNA sequence. 480 bp,  
mRNA, linear, EST 31-OCT-1995

ACCESSION    H73373

VERSION      H73373.1 GI:1047623

KEYWORDS     EST; 5\_prime.

SOURCE       Homo sapiens (human)

ORGANISM     Homo sapiens

REFERENCE    1 (bases 1 to 480)

AUTHORS      Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,  
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.  
and Marra,M.

TITLE        Generation and analysis of 280,000 human expressed sequence tags

JOURNAL      Genome Res. 6 (9), 807-828 (1996)

PUBMED       8889549

COMMENT      Contact: wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Insert Size: 1102

High quality sequence stops: 306

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1102    Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 306.

FEATURES     Location/Qualifiers

source

1..480

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="GDB:3780483"

/db\_xref="taxon:9606"

/clone="IMAGE:229387"

/sex="male"

/dev\_stage="20 week-post conception fetus"

/lab\_host="DH10B (ampicillin resistant)"

/clone\_lib="Soares fetal liver spleen 1NFLS"

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

```

          10      20      30      40      50
ss.H73373      CAGGAACTAGGAGGTTCTCACTGCCCCGAGCAGAGGCCCTACACCCACCGAG
          *  *  *****
ss.DNA57836    GGCACGAGCCAGGAAGTCTCACTGCCCCGAGCAGAGGCCCTACACCCACCGAG
          10      20      30      40      50      60

          60      70      80      90      100     110
ss.H73373      GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG
          *****
ss.DNA57836    GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG
          70      80      90      100     110     120

          120     130     140     150     160     170
ss.H73373      CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTC
          *****
ss.DNA57836    CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTC
          130     140     150     160     170     180

          180     190     200     210     220     230
ss.H73373      CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT
          *****
ss.DNA57836    CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT
          190     200     210     220     230     240

          240     250     260     270     280     290
ss.H73373      ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC
          *****
ss.DNA57836    ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC
          250     260     270     280     290     300

          300     310     320     330     340     350
ss.H73373      CGGCCTCCTTCAACCTCAACGTCACTCAAGTCCAGTCCAGACCTGCTTCACCTXATT
          *****
ss.DNA57836    CGGCCTCCTTCAACCTCAACGTCACTCAAGTCCAGTCCAGACCTGC-TCACCTACTT-
          310     320     330     340     350

          360     370     380     390     400     410
ss.H73373      CTTGCCGGGCGTTCTCCACCTCAGGTGCCCATGTGGGACAGTGCCAGGGTTACAGATGG
          *****
ss.DNA57836    -CTGCCGGGCG-TCCTCCACCTCAGGTGCCCATGT-GGACAGTGCCA-GGCTACAGATGG
          360     370     380     390     400     410

          420     430     440     450     460     470
ss.H73373      CATTGGGGAGCTGTGGTCCAAGCCAGTGTTTTGAGGTGCGGGCCAATTTCATTTTGCAGG
          *  *****
ss.DNA57836    ACT--GGGAGCTGTGGTCCAAGCCAGTG-TCTGAGCTGCGGGCCAATTTCACTCTGCAGG
          420     430     440     450     460     470

          480
ss.H73373      ACAGAGGG
          *****
ss.DNA57836    ACAGAGGGGCGAGGCCCGAGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCACCTA
          480     490     500     510     520     530

          540     550     560     570     580     590
ss.DNA57836    TCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACA
          540     550     560     570     580     590

          600     610     620     630     640     650
ss.DNA57836    GGCAGCCTGCCAATTCTCCTTCTGCGGAGCCAGACATCGGACTGGTTCTGGTGCCAGG
          600     610     620     630     640     650

          660     670     680     690     700     710
ss.DNA57836    CTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCAGGTGGTGACC
          660     670     680     690     700     710
```

ss.DNA57836 AGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACA  
720 730 740 750 760 770

ss.DNA57836 GGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGG  
780 790 800 810 820 830

ss.DNA57836 TCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAG  
840 850 860 870 880 890

ss.DNA57836 AGGACTGCAGGCCATCAGCGTGCACTGTTTCGTATTTGGAGTTCATGCAAAATGAGTGTGT  
900 910 920 930 940 950

ss.DNA57836 TTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAAA  
960 970 980 990

<first sequence: ss.H58326 (length = 418)  
<second sequence: ss.DNA57836 (length = 992)

<400 matches in an overlap of 415: 96.39 percent similarity  
<gaps in first sequence: 1 (1 base), gaps in second sequence: 4 (4 bases)  
<score: 1155 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)  
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:32:19 2005]: 1 sequence found

H58326 yr25c07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone  
IMAGE:206316 5', mRNA sequence. 418 bp,  
mRNA, linear, EST 05-OCT-1995  
ACCESSION H58326  
VERSION H58326.1 GI:1011158  
KEYWORDS EST; WASHU\_est; 5\_prime; merck\_est.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 418)  
AUTHORS Hillier,L., Wilson,R., etal  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT High quality stops: 251; stops: 251; insert: 912.  
FEATURES Location/Qualifiers  
source 1..418  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:3775447"  
/db\_xref="taxon:9606"  
/clone="IMAGE:206316"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares fetal liver spleen 1NFLS"  
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT  
ORIGIN

ss.DNA57836	GGCACGAGCCAGGAAGTCTCACTGCCCCAGCAGAGGCCCTACACCCACCGAG	
	10 20 30 40 50 60	
ss.DNA57836	GCGTGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG	
	70 80 90 100 110 120	
ss.DNA57836	CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTC	
	130 140 150 160 170 180	
ss.DNA57836	CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCAGCCACCACCGCCCATCACCT	
	190 200 210 220 230 240	
ss.H58326		10 20 GAAGGTGGGTGAAGACCCACGAG *****
ss.DNA57836	ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGT-GGTGAAGACCCACGAG	
	250 260 270 280 290	

ss.H58326	30	40	50	60	70	80
	CCGGCCTCCTTCAACCTCAACGTCAACTCAAGTCCAGTCCAGACCTGCTCACCTACTTC					
	*****					
ss.DNA57836	300	310	320	330	340	350
	CCGGCCTCCTTCAACCTCAACGTCAACTCAAGTCCAGTCCAGACCTGCTCACCTACTTC					
ss.H58326	90	100	110	120	130	140
	TGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGG					
	*****					
ss.DNA57836	360	370	380	390	400	410
	TGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGG					
ss.H58326	150	160	170	180	190	200
	GAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGG					
	*****					
ss.DNA57836	420	430	440	450	460	470
	GAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGG					
ss.H58326	210	220	230	240	250	260
	GCAGG-CCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAAC					
	*****					
ss.DNA57836	480	490	500	510	520	530
	GCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAAC					
ss.H58326	270	280	290	300	310	320
	AGCCTGATCGGGAAGGATGGGCAGGTCCACCTXCAGCAGAGACCATGCCACAGGAGGCCT					
	*****					
ss.DNA57836	540	550	560	570	580	590
	AGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCT					
ss.H58326	330	340	350	360	370	380
	GCCAATTGTCCTTCCTTGCCGAGCCAGACATCGGACTGGTTTTTGGTGCCAGGGTGCAA					
	*****					
ss.DNA57836	600	610	620	630	640	650
	GCCAATTCTCCTTCC-TGCCGAGCCAGACATCGGACTGG-TTCTGGTGCCAGGCTGCAA					
ss.H58326	390	400	410			
	ACAACGTCAATGTCCAGXACAGGGXCCTTTAGAXTT					
	*****					
ss.DNA57836	660	670	680	690	700	710
	ACAACGCCAATGTCCAGCACAGCG-CCCTCACAGTGGTGCCCCAGGTGGTGACCAGAAG					
ss.DNA57836	720	730	740	750	760	770
	ATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGC					
ss.DNA57836	780	790	800	810	820	830
	ACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTTCAAGATAGGGAATGGGGAGGTCAGA					
ss.DNA57836	840	850	860	870	880	890
	GGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGAC					
ss.DNA57836	900	910	920	930	940	950
	TGCAGGCCATCAGCGTCACTGTTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAG					
ss.DNA57836	960	970	980	990		
	CTGCTCTTGCCACAAAAA					

<first sequence: ss.R02548 (length = 379)  
<second sequence: ss.DNA57836 (length = 992)

<205 matches in an overlap of 379: 54.09 percent similarity  
<gaps in first sequence: 10 (107 bases), gaps in second sequence: 5 (14 bases)  
<score: 443 (match = 3, mismatch = 0, gap penalty = 8 + 1 per base)  
<endgaps not penalized

GenBank (Release 149, aug 2005)[Sep 20 13:31:10 2005]: 1 sequence found

R02548        ye80a07.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone  
IMAGE:124020 3', mRNA sequence. 379 bp,  
mRNA, linear, EST 31-MAR-1995  
ACCESSION    R02548  
VERSION      R02548.1 GI:752284  
KEYWORDS     EST; WASHU\_est; 3\_prime; merck\_est.  
SOURCE       Homo sapiens (human)  
ORGANISM     Homo sapiens  
REFERENCE    1 (bases 1 to 379)  
AUTHORS      Hillier,L., Wilson,R., etal  
TITLE        The WashU-Merck EST Project  
JOURNAL      Unpublished (1995)  
COMMENT      High quality stops: 309; stops: 309; insert: 846.  
FEATURES     Location/Qualifiers  
              source        1..379  
                             /organism="Homo sapiens"  
                             /mol\_type="mRNA"  
                             /db\_xref="GDB:476565"  
                             /db\_xref="taxon:9606"  
                             /clone="IMAGE:124020"  
                             /sex="male"  
                             /dev\_stage="20 week-post conception fetus"  
                             /lab\_host="DH10B (ampicillin resistant)"  
                             /clone\_lib="Soares fetal liver spleen 1NFLS"  
                             /note="Organ: Liver and spleen; Vector: pT7T3D (Pharmacia)  
                             with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
                             1st strand cDNA was primed with a Pac I - oligo(dT) primer  
                             [5' AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3']  
                             double-stranded cDNA was ligated to Eco RI adaptors  
                             (Pharmacia), digested with Pac I and cloned into the Pac I  
                             and Eco RI sites of the modified pT7T3 vector. Library  
                             went through one round of normalization. Library  
                             constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

ss.DNA57836	GGCACGAGCCAGGA	ACTAGGAGGTTCT	CACTGCCCCGAGC	AGAGGCCCTACAC	CCACCGAG	
	10	20	30	40	50	60
ss.DNA57836	GCATGGGGCTCCCT	GGGCTGTTCTGCT	TGGCCGTGCTGGCT	GCCAGCAGCTTCT	CCAAGG	
	70	80	90	100	110	120
ss.DNA57836	CACGGGAGGAAGAA	ATTACCCCTGTGGT	CTCCATTGCCTACA	AAGTCCTGGAAGT	TTTTCC	
	130	140	150	160	170	180
ss.DNA57836	CCAAAGGCCGCTGG	GTGCTCATAACCTG	CTGTGCACCCCAGC	CCACCACCGCCCAT	CACCT	
	190	200	210	220	230	240
ss.R02548					10	
					TTTTTGTGGCAAGAGC	
					* * * * *	
ss.DNA57836	ATTCCTCTGTGGA	ACCAAGAACATCA	AGGTGGCCAAGAAG	GGTGGTGAAGACCC	ACGAGC	
	250	260	270	280	290	300

ss.R02548  
AG--CTAAAACACTCA--TTTTGCATGAACTCCAAATACGAACAGTGCACGCTGAT-G  
\* \* \* \* \*  
ss.DNA57836  
CGGCCTCCTTCAACCTCAACGTCACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT  
310 320 330 340 350 360

ss.R02548  
GCCTGCAGTCCTCTGCCGTGCTTGCTCTCTGGACAGT-TCATTCTACATGGCTGCTGCT  
\*\*\* \* \*\*\*\*\* \*\* \*\* \* \* \*\*\*\*\* \*\* \*\*\*\*\* \* \*\*\*  
ss.DNA57836  
GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACA-----GATGCA  
370 380 390 400 410

ss.R02548  
TTGCGTCCTCT-GACCTCCCCATTCCCTATCCT--GAACCCCCCAAACCTCTTCACTC  
\*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*  
ss.DNA57836  
CTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAG  
420 430 440 450 460 470

ss.R02548  
190 200 210 220  
AGACGGCGGGTGCTCCTGTAGAGCGGCAAGGCAAGG-----  
\*\* \*\*\* \*\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*  
ss.DNA57836  
AG-GGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCTCGGGCAGCCCACCTATCA  
480 490 500 510 520 530

ss.R02548  
-----  
ss.DNA57836  
CCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGC  
540 550 560 570 580 590

ss.R02548  
-----ATGGGGCTCTCCAGGGGACCTGCCAGTCCCTCCATCTTCTGGTCAACACCT  
\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*  
ss.DNA57836  
AGCCTGCCAATTCTCCTTCTGCCGAGCCAGACA-TCGGACTGGTTCTGGTGCCAGGCT  
600 610 620 630 640 650

ss.R02548  
280 290 300 310 320 330  
GGGCCCA--CGTGGGTCCAGCCAGCATCCTGGGAGGTGATGGCCGAGTGGGAGGCAA  
\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*  
ss.DNA57836  
GCAAACAACGCCAATGTCCAGCACAGCGCCCTCACA-GTGGTGCCCCCAGGTGGTGACCA  
660 670 680 690 700 710

ss.R02548  
340 350 360 370  
GGXTGCCAACAGCACGA-TGGTGGGGTGCCTGGGGAAACCC-TCCTGG  
\* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \* \*  
ss.DNA57836  
G-----AAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCTTGCCGCT  
720 730 740 750 760

ss.DNA57836  
CTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTTCCAGGATAGGGAATGG  
770 780 790 800 810 820

ss.DNA57836  
GGAGGTGAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCAC  
830 840 850 860 870 880

ss.DNA57836  
GGCAGAGGACTGCAGGCCATCAGCGTGCCTGTTTCGTATTTGGAGTTCATGCAAAATGAG  
890 900 910 920 930 940

ss.DNA57836  
TGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAA  
950 960 970 980 990

## Genetic Instability in Epithelial Tissues at Risk for Cancer

WALTER N. HITTELMAN

Department of Experimental Therapeutics, The University of Texas  
M. D. Anderson Cancer Center, Houston, Texas 77030, USA

**ABSTRACT:** Epithelial tumors develop through a multistep process driven by genomic instability frequently associated with etiologic agents such as prolonged tobacco smoke exposure or human papilloma virus (HPV) infection. The purpose of the studies reported here was to examine the nature of genomic instability in epithelial tissues at cancer risk in order to identify tissue genetic biomarkers that might be used to assess an individual's cancer risk and response to chemopreventive intervention. As part of several chemoprevention trials, biopsies were obtained from risk tissues (i.e., bronchial biopsies from chronic smokers, oral or laryngeal biopsies from individuals with premalignancy) and examined for chromosome instability using *in situ* hybridization. Nearly all biopsy specimens show evidence for chromosome instability throughout the exposed tissue. Increased chromosome instability was observed with histologic progression in the normal to tumor transition of head and neck squamous cell carcinomas. Chromosome instability was also seen in premalignant head and neck lesions, and high levels were associated with subsequent tumor development. In bronchial biopsies of current smokers, the level of ongoing chromosome instability correlated with smoking intensity (e.g., packs/day), whereas the chromosome index (average number of chromosome copies per cell) correlated with cumulative tobacco exposure (i.e., pack-years). Spatial chromosome analyses of the epithelium demonstrated multifocal clonal outgrowths. In former smokers, random chromosome instability was reduced; however, clonal populations appeared to persist for many years, perhaps accounting for continued lung cancer risk following smoking cessation.

**KEYWORDS:** chromosome instability; epithelial cells; aerodigestive tract; chemoprevention; cancer risk

### THE NEED FOR BIOMARKERS OF CANCER RISK AND RESPONSE TO INTERVENTION

Epithelial cancers remain a major health challenge in the world. Despite improvements in staging and the application and integration of surgery, radiotherapy, and chemotherapy, the 5-year survival rate for individuals with lung cancer is only about 15%.<sup>1</sup> Even if strategies for early detection are successful and lung cancers are detected at a stage where local tumor resection and treatment is curative, these patients will still be at significant risk for developing second primary tumors

Address for correspondence: Dr. Walter N. Hittelman, Department of Experimental Therapeutics, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd. (Box 19), Houston, Texas 77030. Voice: 713-792-2961; fax: 713-792-3754.  
whittelm@mdanderson.org



associated with the problem of field cancerization.<sup>2</sup> Similarly, for individuals with a first head and neck primary tumor, even if the first malignancy is successfully treated, the risk of developing a second primary in the tobacco smoke-exposed field is approximately 40%.<sup>3</sup> Similar cancer risk estimates exist for individuals who exhibit severe dysplasia in premalignant epithelial lesions.<sup>4</sup> For these reasons, it is important to focus on chemopreventive strategies to prevent the development of epithelial malignancies.

Several problems confront chemoprevention trials designed to identify efficacious agents.<sup>5</sup> First, chemoprevention trials with cancer incidence as a primary endpoint require tens of thousands of subjects and tens of years of intervention and follow-up for statistical evaluation. For example, a recently reported trial involved 30,000 subjects and required 10 years in order to examine the impact of prevention strategies on lung cancer development, only to find a possible increased lung cancer incidence in current smokers who received  $\beta$ -carotene.<sup>6</sup>

The problem of large, long-term trials results from the difficulty in identifying individuals at highest cancer risk who might best benefit from chemopreventive intervention. For example, 20 pack-year smokers, while known to be at relatively increased risk for developing lung cancer, have approximately a 10% lifetime risk for developing lung cancer.<sup>7</sup> This seriously limits the number of potentially useful strategies that can be clinically explored. A second problem facing chemoprevention trials is that little is known about what agents are likely to have efficacy, and even less is known regarding proper doses, schedules, and durations of treatment. Part of the reason for this problem is that too little is known about the physiologic processes that drive epithelial cancer development.

In order to reduce the number of subjects and the time required to carry out chemoprevention trials and thus allow the exploration of multiple prevention strategies, two types of advances are necessary. First, it is important to identify individuals at significantly increased cancer risk who might best benefit from different types of intervention. Second, in order to allow the rapid identification of agents, doses, and schedules of potentially efficacious agents, it is necessary to identify and validate surrogate endpoints of response that indicate whether the agents are having a positive impact on the target tissue during the chemopreventive intervention.

One approach to identifying individuals at increased aerodigestive tract cancer risk is to explore epidemiologic features of potential subjects. Molecular epidemiologic studies are beginning to identify intrinsic host factors that place some individuals at increased cancer risk, especially those with a chronic smoking history.<sup>8</sup> Most intrinsic factors identified thus far reflect levels of carcinogen metabolism, repair capabilities of the host following DNA damage, and other measures of intrinsic cellular sensitivity to mutagens. While these factors can provide statistically significant risk ratios in case-control studies that are controlled for tobacco exposure, the detected risk ratios usually fall in the range of 1.5 to 10. Unfortunately, this is not sufficient for the individualization of treatment and is not sufficiently high to significantly reduce the numbers of subjects required for chemoprevention trials with cancer incidence as the primary endpoint.

Another approach to identifying individuals at increased cancer risk is to directly examine the target tissue of individuals with known carcinogen exposure (e.g., chronic tobacco smoke exposure), who have evidence of target organ dysfunction

(e.g., chronic obstructive pulmonary disease, changes in voice quality), or who have clinical evidence of premalignancy (e.g., bronchial metaplasia/dysplasia, oral leukoplakia/erythroplakia, cervical intraepithelial neoplasia). The conventional standard for assessing cancer risk in these situations is the degree of histological change. However, while individuals who show moderate to severe dysplasia are known to be at increased cancer risk when compared to individuals with lesser histologic changes, it is often difficult to distinguish reactive changes to carcinogenic insult from initiated and progressing lesions. Similarly, upon cessation of carcinogenic insult, histologic changes may reverse yet cancer risk may continue for many years. For example, while smoking cessation is associated with decreased bronchial metaplasia,<sup>9</sup> increased lung cancer risk continues for many years beyond smoking cessation.<sup>10</sup> In fact, nearly half the newly diagnosed lung cancer cases in the USA occur in former smokers.<sup>11</sup>

The development of assays to identify individuals at high epithelial cancer risk and to directly assess response to intervention in the target tissue is therefore an important research goal. Such assays should be objective and easily quantifiable and, if possible, minimally invasive. Moreover, they should reflect both the disease process and the targeted pathway and thereby be useful in assessing risk and monitoring response to intervention as well as directly testing the hypothesized mechanism of action of the chemopreventive strategy.

In the chemoprevention setting it is important to recognize that one does not know the location of the future cancer. Thus, assays must necessarily be carried out on random biopsies of the field at risk. Even if there are clinically evident premalignant lesions, this does not mean that this is the likely site for a future malignancy. For example, nearly half of the cancers that develop in individuals with oral leukoplakia arise away from the original index lesion. Similarly, since many newly diagnosed lung cancers arise in the peripheral parts of the lung (e.g., adenocarcinomas), especially in former smokers, and since endobronchoscopy predominantly accesses central components of the lung, it is important to identify biomarkers that can reflect global processes ongoing in the target epithelial field associated with increased cancer risk. Their discovery requires a better understanding of the tumorigenesis process in epithelial fields at cancer risk.

#### THE RATIONALE FOR STUDYING GENOMIC INSTABILITY AS A MARKER OF RISK

Tumors of the aerodigestive tract have been proposed to reflect a "field cancerization" process whereby the whole tissue is exposed to carcinogenic insult (e.g., tobacco smoke) and is at increased risk for multistep tumor development.<sup>12,13</sup> Several types of clinical and laboratory data support this notion, including the frequent occurrence of synchronous primary and subsequent second primary tumors in the aerodigestive tract (frequently exhibiting dissimilar histologies as well as distinct genetic signatures<sup>14-16</sup>) and the presence of premalignant lesions that precede and/or accompany the tumor in the exposed tissue field.<sup>17</sup> The notion of a multistep tumorigenesis process is further supported by serial clinical and histologic evaluations of

target tissue or exfoliated cells where increasing degrees of histological abnormalities are observed over time.<sup>18</sup>

A working model for aerodigestive tract tumorigenesis is illustrated in FIGURE 1. Tumorigenesis in the face of carcinogenic exposure likely involves a chronic process of tissue injury and wound healing. DNA damage induced by the carcinogen is likely fixed into permanent genetic changes (e.g., chromosome damage, chromosome non-disjunction, gene mutation, gene deletion, etc.) during the process of proliferation. This damage would be expected to be distributed throughout the exposed tissue field leading to a background of generalized genomic damage (depicted in FIGURE 1 as a background mat of increasing density). Chronic injury and repair likely leads to the accumulation of cells with increasing amounts of genetic changes as well as the outgrowth of abnormal clones (triangles in FIGURE 1) carrying an accumulation of genetic changes important for selective survival, dysregulated growth, and preferential epithelial take-over by initiated clones (see FIGURE 2).

Cellular and molecular evidence for the field carcinogenesis and multistep tumorigenesis model comes from many laboratories.<sup>19,20</sup> With the advent of a wide array of molecular technologies, a large number of specific molecular genetic and epigenetic changes involving specific oncogenes, tumor suppressor genes, cell regulatory genes, and repair genes have now been described for aerodigestive tract cancers. The identification of these specific molecular changes have now provided probes to explore specific events occurring in premalignant lesions adjacent to aerodigestive tract tumors.<sup>21-24</sup> Frequently, these premalignant lesions showed a subset of the same molecular changes found in the associated tumor, suggesting that these lesions might represent precursor lesions for the associated tumors (i.e., a manifestation of

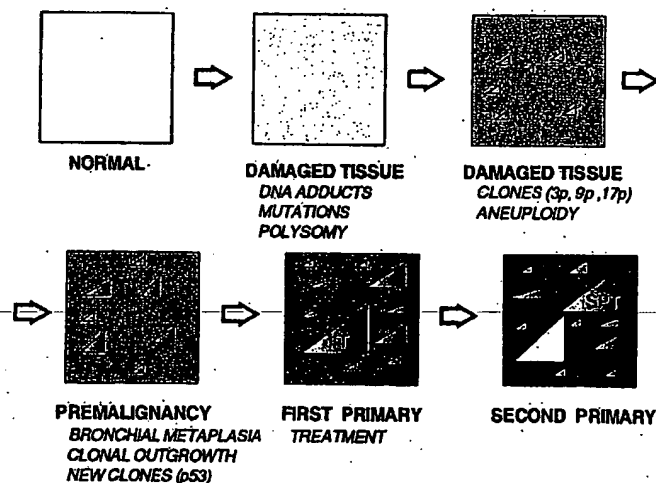
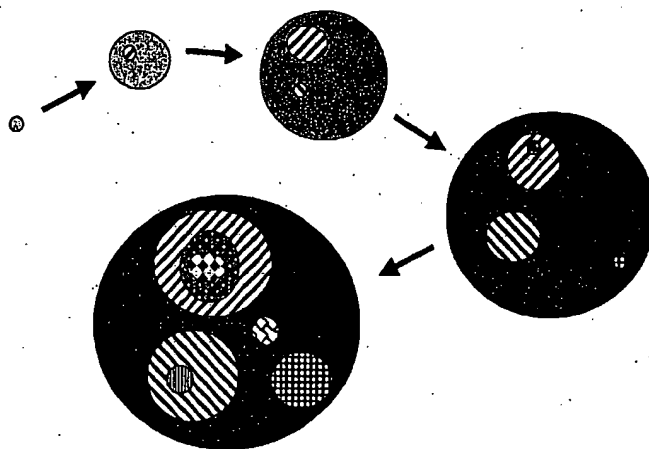


FIGURE 1. Field cancerization and multistep tumorigenesis.



**FIGURE 2.** Multiple focal clonal evolution during multistep tumorigenesis.

a multistep tumorigenesis process). For example, studies of the premalignant lesions adjacent to head and neck tumors have provided evidence for a gradual accumulation of genetic alterations accompanied by evidence for dysregulation of cellular control mechanisms (e.g., alterations in expression of PCNA, EGFR, TGF- $\beta$ , p53, and cyclin D1).<sup>25-28</sup>

These types of studies have now also been applied to the target epithelium of individuals at increased risk for aerodigestive tract cancer (i.e., individuals with a chronic smoking/alcohol history and/or prior aerodigestive tract cancer). Several groups (using polymerase chain reaction, PCR, analysis of microdissected epithelium) have now demonstrated the presence of clonal outgrowths in the target premalignant epithelium of individuals at increased risk for cancer.<sup>29-31</sup> For example, examination of bronchial biopsies derived from individuals with a 20 pack-year smoking history demonstrated that 76% of the cases showed evidence for LOH (3p14, 9p21, or 17p13) in at least one of six lung biopsy sites. On a per site basis, some form of LOH was observed in 25% of the sites examined.<sup>29</sup>

If aerodigestive tract cancer development reflects a field cancerization process involving multistep events, then risk and response information should be able to be derived from random biopsies or exfoliated cells from the field at risk or from assessments of tissue undergoing similar processes. Hypothetically, lesions exhibiting the greatest degree of genomic instability, clonal outgrowth, and abnormal epithelial regulation would be at the highest relative aerodigestive tract cancer risk. Similarly, an active chemopreventive intervention might be expected to decrease these manifestations of risk. Reduced risk manifestations include decreased levels of ongoing genetic instability, decreased frequency of clonal outgrowths, and increased epithelial growth regulation.

### THE MEASUREMENT OF CHROMOSOME INSTABILITY USING CHROMOSOME *IN SITU* HYBRIDIZATION

Molecular genetic techniques, while extremely useful for detecting clonal changes in target tissues, are somewhat limited in their ability to detect random genetic instability. Conventional cytogenetic assays are useful for detecting chromosome instability and clonal chromosome changes. However, they require numbers of dividing cells for karyotypic analysis that are difficult to attain in the setting of biopsies acquired during the course of a chemoprevention trial. A technique was therefore needed that would allow chromosome instability measurements in situations where few cells are available (e.g. small biopsies, brushings, or sputum samples) and where the target material might be fixed. It was also desirable to have a technique that would be adaptable to tissue sections, whereby spatial information could be retained and genotype/phenotype associations could be determined on the same or adjacent tissue sections. The technique of *in situ* hybridization (ISH) involves the use of DNA probes that recognize either chromosome-specific repetitive target sequences, chromosome single gene copy sequences, or sequences along the whole chromosome length or chromosome segments.<sup>32</sup> We have adapted the ISH technique for formalin-fixed, paraffin-embedded tissue sections and have applied it to a variety of tissues, including the aerodigestive tract.<sup>33,34</sup>

Using probes that label the centromere regions of specific chromosomes, this assay permits determination of the average chromosome number per cell for each specimen. This assay is also useful for detecting generalized chromosome instability during the tumorigenesis process. Normal diploid populations should have two copies of each autosomal chromosome and should rarely show three or more chromosome copies per cell (chromosome polysomy), especially in tissue sections where nuclear truncation results in an under-representation of chromosome copy number. Thus, the detection of cells with three or more chromosome copies would indicate the presence of chromosome instability.

To examine this technique's potential for characterizing the multistep tumorigenesis process in the aerodigestive tract, we measured the fraction of cells exhibiting three or more chromosome copies in apparently contiguous epithelial transitions from normal to hyperplastic to dysplastic to carcinomas, all on a single tissue slice of head and neck squamous cell carcinomas.<sup>34</sup> In these specimens, greater than 35% of the cases of adjacent "normal" epithelium, greater than 65% of the cases of hyperplastic epithelium, and greater than 95% of the dysplastic and tumor regions showed evidence of chromosome polysomy. Of interest, similar transitions of chromosome instability were observed with at least four different chromosome probes. Similar trends have also been observed in amenable tissue from other epithelial malignancies, including cervix, bladder, and breast.<sup>35</sup> These results thus suggested that the notions of field cancerization and multistep tumorigenesis might apply to several epithelial tissues and that measures of chromosome instability might be useful for monitoring this process.

In the situations described above, the premalignant lesions examined might be considered to represent epithelium at 100% risk of being in a cancer field, since they were located in the adjacent epithelium to the cancer. This then raises the question of the nature of genetic instability in the epithelium of individuals at increased risk

for developing cancer. To explore this issue, we obtained biopsies during the course of leukoplakia chemoprevention trials exploring the use of 13-*cis*-retinoic acid in reversing leukoplakia and probed them for genetic instability using *in situ* hybridization. In one retrospective study and in one prospective study of subjects with oral leukoplakia, the results indicate that those subjects whose pretreatment biopsies harbor relatively high levels of genomic instability (i.e., more than 3% of the cells examined showing at least 3 chromosome 9 copies per cell) have a significantly higher likelihood of suffering early onset of head and neck cancer.<sup>36,37</sup> Interestingly, half of the tumors that did develop occurred away from the biopsy site used to measure genetic instability. This result suggests that genomic instability measurements in carcinogen-exposed tissue can provide useful cancer risk estimates.

#### THE RELATIONSHIP BETWEEN TOBACCO EXPOSURE AND CHROMOSOME INSTABILITY

In recent years, the aerodigestive tract chemoprevention group at M.D. Anderson Cancer Center has initiated three sequential biomarker-associated chemoprevention trials involving chronic smokers with a greater than 20 pack-year smoking history. In each of these studies, endobronchial biopsies were obtained from six defined sites within the lung, including the carina and at bifurcation points at the upper, middle, and lower right lung and at the upper and lower left lung. Biopsies were obtained prior to and following chemopreventive intervention and were subjected to *in situ* hybridization analysis in addition to analyses for other biomarkers. The first important finding was that some degree of chromosome polysomy was evident in all lung sites examined, and this was observed independently of the particular chromosome probe utilized.<sup>38</sup> This finding supports the notion that random chromosome changes may be occurring throughout the exposed lung field.

In a second study, bronchial biopsies were obtained from individuals with a 20 pack-year smoking history. In this study, most of the subjects involved were current smokers.<sup>39</sup> Interestingly, all cases who showed metaplasia at one of six biopsy sites also showed chromosome polysomy in at least one biopsy site; overall, 88% of the sites showed some evidence of chromosome 9 polysomy.<sup>40</sup> Evidence for genetic instability was also detected in patients who did not show evidence of bronchial metaplasia in any of six biopsy sites despite a strong smoking history. In fact, more than 90% of the cases and more than 60% of the sites showed significant chromosome polysomy (i.e., at least three copies in at least 2 % of the cells examined). These results suggest that the lungs of long-term smokers show significant evidence of genetic instability, and this instability can be detected throughout the accessible bronchial tree, even when bronchial metaplasia is not evident.

These studies in current smokers has allowed us to examine the relationship between the levels of genetic instability detected and subject characteristics such as smoking status (current or former), smoking history, and lung tissue pathologic changes. Evaluable biopsy material has now been obtained from more than 108 current smokers, including more than 480 evaluable biopsy sites. The mean metaplasia index in these current smokers was 30.4%. For the total population studied, the median chromosome index for the bronchial biopsies was 1.41 (range, 1.04-1.61)

and the median chromosome polysomy index was 2.0% (range 0–8.7%). This can be compared to a mean chromosome index between 1.2–1.4 for lymphocytes and very rare chromosome polysomy. Interestingly, the intrasubject variability in chromosome instability was relatively low in most subjects and was less than the intersubject variability. These results suggested that chronic smokers harbor detectable chromosome instability throughout the accessible bronchial tree (supporting the field carcinogenesis notion) and that information from one biopsy site might yield representative information for the rest of the lung field.

Since most of the current smokers exhibited bronchial metaplasia in at least one of the biopsied sites, this allowed us to examine the relationship between chromosome instability and histologic changes, both on a site-by-site basis and on a per case basis. On a site-by-site basis, the chromosome indices of lesions showing squamous metaplasia were similar to those not showing metaplasia (i.e., median 1.43 vs. 1.43), and the degree of chromosome polysomy in metaplastic lesions were only slightly higher than in non-metaplastic sites (medians: 2.2% vs. 1.8%, respectively). Thus, the presence or absence of squamous metaplasia at a biopsy site does not necessarily correlate with the degree of underlying genomic instability. On the other hand, those subjects with metaplasia indices of at least 15% also showed higher levels of chromosome polysomy than did subjects with metaplasia index below 15% (medians: 2.4% vs. 1.8%,  $p = 0.005$ ). Thus, these chromosome instability assessments in current smokers appeared to reflect a more global process in the lung field.

Tobacco exposure has been shown to significantly increase the risk of developing lung cancer, and the degree of risk is related to the extent of tobacco exposure. We were interested in determining the relationship between individuals' smoking history parameters and the levels of chromosome change found in their lungs following years of tobacco exposure. While there was significant intersubject variation for similar tobacco exposure histories, overall there was a significant correlation between the degree of chromosome polysomy and the intensity of ongoing tobacco exposure (packs/day,  $p = 0.02$  on a per site basis) and with the extent of tobacco exposure (pack-years,  $p = 0.003$ ). Thus the amount of chromosome polysomy reflects the intensity and extent of tobacco exposure. At the same time, individuals with similar smoking histories showed widely divergent amounts of chromosome polysomy, possibly reflecting differences in intrinsic sensitivity between subjects. There was also strong correlation between the chromosome index and the duration of the smoking history (smoking years) and total accumulated exposure (pack-years,  $p = 0.0001$ ). These results suggest that tobacco exposure is associated with the initiation and accumulation of chromosome instability in the exposed lung; however individuals are differentially sensitive to carcinogenic insult. The working hypothesis is that those individuals who accumulate the highest degree of chromosome changes will be at the highest lung cancer risk.

Many of the bronchial biopsies from chronic smokers examined by *in situ* hybridization showed a rise in the chromosome index above that expected for a diploid cell population, especially in subjects with an extensive smoking history. The rise in chromosome index was also accompanied by an increase in the fraction of cells exhibiting at least 3 chromosome copies per cell. To determine if a rise in the tissue chromosome index was due to clonal expansion of populations with chromosome trisomy, the chromosome copy number and relative coordinates of each cell scored in

the bronchial epithelium was recorded and a spatial genetic map was created.<sup>41</sup> We then developed algorithms for calculating localized chromosome indices within the tissue. Since trisomic clones would have, on average, three chromosomes instead of two, those cells involved in neighborhoods with chromosome indices three-halves that of diploid populations could be marked as being part of a trisomic clone. Similarly, groups of cells with chromosome indices half that of diploid populations could be marked as being part of a monosomic clone. This allowed the generation of a second-order, two-dimensional genetic map representation of the bronchial epithelium showing the relative locations of cells involved in monosomic and trisomic clonal outgrowths. When adjacent tissue sections from the same bronchial biopsy were probed separately for different chromosomes, the detected clones appeared to occupy separate subregions of the epithelium. This result suggests that not only are the lungs of chronic smokers undergoing a process of genetic instability, they are experiencing the outgrowth of multiple clones throughout the exposed lung field, as postulated by the models shown in FIGURES 1 and 2. One advantage of this clonal approach is that the contribution of both monosomic and multisomic clones can be detected.

Since smoking cessation has been suggested to reduce the lung cancer risk, it was of interest to determine whether the levels of chromosome instability would decrease following smoking cessation. This question was possible to examine because our third sequential chemoprevention trial involved subjects who had discontinued smoking. So far, more than 220 subjects (more than 650 biopsies) who have quit smoking (mean 9.9 quit-years) have been evaluated for chromosome instability in their lungs. Despite the fact that the mean metaplasia index in this group is 5.8% (considerably less than that in current smokers), chromosome instability is still observed in the majority of subjects.<sup>42</sup> While the mean chromosome polysomy level is reduced to 1.0%, some individuals continue to show polysomy levels above 5%. Interestingly, while the overall chromosome polysomy levels were reduced in these individuals who stopped smoking, the mean chromosome index remained at about 1.4 with some individuals exhibiting chromosome indices as high as 1.8. Initial chromosome mapping studies suggest that while random chromosome instability seems to decrease following smoking cessation, the clonal outgrowths may remain for many years in the lung. The working hypothesis is that those individuals who show the greatest degree of remaining chromosome instability are at the highest lung cancer risk despite smoking cessation. Long-term follow-up on these subjects will be necessary to test this hypothesis.

---

## SUMMARY AND CONCLUSIONS

Aerodigestive tract tumorigenesis appears to be a multistep process taking place throughout the tissue fields of exposure. When viewed in the context of chromosome changes, carcinogen exposure appears to be associated with the random acquisition of chromosome polysomy throughout the exposed field, the degree of which is related to the degree and extent of carcinogen exposure as well as to the intrinsic susceptibility of the exposed individual. Continued exposure leads to continued acquisition of new changes and, in association with chronic wound-healing processes, to the



accumulation of clonal outgrowths throughout the target tissue. Although the ultimate malignancy may occur in only one or few tissue sites, manifestations of the instability process that drives tumorigenesis is globally present in the tissue. Thus random biopsies may provide useful risk information for the exposed field as a whole. Even when carcinogen exposure is reduced or chemopreventive strategies are initiated and histologic manifestations of the tumorigenesis process subside, the genetic scars of prior exposure remain in the form of clonal outgrowths and may explain continued lung cancer risk in ex-smokers. Future chemoprevention strategies need to focus on reducing the degree of chromosome instability and on trying to eliminate residual abnormal clonal outgrowths in the aerodigestive tract. In this setting, the measurement of chromosome instability in the target tissue will be useful in assessing cancer risk as well as response to intervention.

#### ACKNOWLEDGMENTS

The studies reviewed here represent one component of the collaborative efforts of the Aerodigestive Tract Chemoprevention team at The University of Texas M.D. Anderson Cancer Center, Houston, Texas. The studies were supported in part by National Institutes of Health-National Cancer Institute Grants CA-52051, CA-68437, CA 79437, CA 16672, CA 68089, CN 25433, CA 86390, CA 70907, NIH DE 13157, and the State of Texas Tobacco Research Fund.

#### REFERENCES

1. LANDIS, S.H., T. MURRAY, S. BOLDEN & P.A. WINGO. 1998. Cancer statistics, 1998. *CA Cancer J. Clin.* 48: 6-29.
2. JOHNSON, B.E. 1998. Second lung cancers in patients after treatment for an initial lung cancer. *J. Natl. Cancer Inst.* 90: 1335-1345.
3. LIPPMAN, S.M. & W.K. HONG. 1989. Second malignant tumors in head and neck squamous cell carcinoma: The overshadowing threat for patients with early stage of disease. *Int. J. Radiat. Oncol. Biol. Phys.* 17: 691-694.
4. SILVERMAN, S.J., JR., M. GORSKY & F. LOZADA. 1984. Oral leukoplakia and malignant transformation: a follow-up study of 257 patients. *Cancer* 53: 563-568.
5. LIPPMAN, S.M., J.S. LEE, R. LOTAN, *et al.* 1990. Biomarkers as intermediate endpoints in chemoprevention trials. *J. Natl. Cancer Inst.* 82: 555-560.
6. HEINONEN, O.P., D. ALBANES & THE ALPHA-TOCOPHEROL, BETA CAROTENE CANCER PREVENTION STUDY GROUP. 1994. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N. Engl. J. Med.* 330: 1029-1035.
7. PETO, R., S. DARBY, H. DEO, *et al.* 2000. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *Brit. Med. J.* 321: 323-329.
8. PERERA, F.P. 1996 Molecular epidemiology: insights into cancer susceptibility, risk assessment, and prevention. *J. Natl. Cancer Inst.* 88: 496-509.
9. LEE, J.S., S.M. LIPPMAN, S.E. BENNER, *et al.* 1994. Randomized placebo-controlled trial of isotretinoin in chemoprevention of bronchial squamous metaplasia. *J. Clin. Oncol.* 12: 937-941.

10. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. 1990. The health benefits of smoking cessation: a report of the Surgeon General. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. DHHS Pub. No. (CDC) 90-8416.
11. TONG, L., M.R. SPITZ, J.J. FAEGER, *et al.* 1996. Lung cancer in former smokers. *Cancer* 78: 1004-1010.
12. SLAUGHTER, D.P., H.W. SOUTHWICK & W. SMEJKAL. 1953. Field cancerization in oral stratified squamous epithelium: clinical implications of multicentric origin. *Cancer* 6: 963-968.
13. FARBER, E. 1984. The multistep nature of cancer development. *Cancer Res.* 44: 4217-4223.
14. CHUNG, K.Y., T. MUKHOPADHYAY, J. KIM, *et al.* 1993. Discordant p53 gene mutations in primary head and neck cancers and corresponding second primary cancers of the upper aerodigestive tract. *Cancer Res.* 53: 1676-1683.
15. SCHOLZ, A.G.M., J.A. WOOLGAR, M.A. BOYLE, *et al.* 1998. Synchronous oral carcinomas: independent or common clonal origin? *Cancer Res.* 58: 2003-2006.
16. GLUCKMAN, J.O., J.D. CRISMAN & J.O. DONEGAN. 1980. Multicentric squamous cell carcinoma of the upper aerodigestive tract. *Head Neck Surg.* 3: 90-96.
17. AUERBACH, O., A.P. STOUT, E.C. HAMMOND, *et al.* 1961. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N. Engl. J. Med.* 265: 253-267.
18. SACCOMANNO, G., V.E. ARCHER, O. AUERBACH, *et al.* 1974. Development of carcinoma of the lung as reflected in exfoliated cells. *Cancer* 33: 256-270.
19. IZZO, J.G. & W.N. HITTELMAN. 1999. Characterization of multistep tumorigenesis by in situ hybridization. In *Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications*. M. Andreeff & D. Pinkel, Eds.: 173-208. John Wiley & Sons, Inc. New York.
20. HITTELMAN, W.N. 1999. Molecular cytogenetic evidence for multistep tumorigenesis: implications for risk assessment and early detection. In *Molecular Pathology of Cancer*. S. Srivastava, D.E. Hensen & A. Gazdar, Eds.: 385-404. IOS Press. Amsterdam, The Netherlands.
21. SUNDARESAN, V., P. GANLY, R. HASLETON, *et al.* 1992. p53 and chromosome 3 abnormalities, characteristic of malignant lung tumours, are detectable in preinvasive lesions of the bronchus. *Oncogene* 7: 1989-1997.
22. KISHIMOTO, Y., K. SUGIO, J.Y. HUNG, *et al.* 1995. Allele-specific loss in chromosome 9p loci in preneoplastic lesions accompanying non-small-cell lung cancers. *J. Natl. Cancer Inst.* 87: 1224-1229.
23. CALIFANO, J., P. VAN DER RIET, W. WESTRA, *et al.* 1996. Genetic progression model for head and neck cancer: implications for field cancerization. *Cancer Res.* 56: 2488-2492.
24. PARK I.W., I.I. WISTUBA, A. MAITRA, *et al.* 1999. Multiple clonal abnormalities in the bronchial epithelium of patients with lung cancer. *J. Natl. Cancer Inst.* 91: 1863-1868.
25. SHIN, D.M., N. VORAVUD, J.Y. RO, *et al.* 1994. Sequential increases in proliferating cell nuclear antigen expression in head and neck tumorigenesis: a potential biomarker. *J. Natl. Cancer Inst.* 85: 971-978.
26. SHIN, D.M., J.Y. RO, W.K. HONG, *et al.* 1994. Dysregulation of epidermal growth factor receptor expression in premalignant lesions during head and neck tumorigenesis. *Cancer Res.* 54: 3153-3159.
27. SHIN, D.M., J. KIM, J.Y. RO, *et al.* 1994. Activation of p53 gene expression in premalignant lesions during head and neck tumorigenesis. *Cancer Res.* 54: 321-326.
28. IZZO, J.G., V.A. PAPADIMITRAKOPOULOU, X.Q. LI, *et al.* 1998. Dysregulated cyclin D1 expression early in head and neck tumorigenesis: in vivo evidence for an association with subsequent gene amplification. *Oncogene* 17: 2313-2322.
29. MAO, L., J.S. LEB, J.M. KURIE, *et al.* 1997. Clonal genetic alterations in the lungs of current and former smokers. *J. Natl. Cancer Inst.* 89: 857-862.

30. WISTUBA, I.I., S. LAM, C. BEHRENS, *et al.* 1997. Molecular damage in the bronchial epithelium of current and former smokers. *J. Natl. Cancer Inst.* 89: 1366-1373.
31. MAO, L., J.S. LEE, Y.H. FAN, *et al.* 1996. Frequent microsatellite alterations at chromosomes 9p21 and 3p14 in oral premalignant lesions and their value in cancer risk assessment. *Nature Med.* 2: 682-685.
32. PODDIGHE, P.J., F.C. RAMAKERS & A.H. HOPMAN. 1992. Interphase cytogenetics of tumours. *J. Pathol.* 166: 215-224.
33. KIM, S.Y., J.S. LEE, J.Y. RO, *et al.* 1993. Interphase cytogenetics in paraffin sections of lung tumors by non-isotopic *in situ* hybridization. Mapping genotype/phenotype heterogeneity. *Am. J. Pathol.* 142: 307-317.
34. VORAVUD, N., D.M. SHIN, J.Y. RO, *et al.* 1993. Increased polysomies of chromosomes 7 and 17 during head and neck multistage tumorigenesis. *Cancer Res.* 53: 2874-2883.
35. HITTELMAN, W.N. 1999. Genetic instability assessments in the lung cancerization field. *In Lung Tumors: Fundamental Biology and Clinical Management.* C. Brambilla & E. Brambilla, Eds.: 255-267. Marcel Dekker. New York.
36. LEE, J.S., S.Y. KIM, W.K. HONG, *et al.* 1993. Detection of chromosomal polysomy in oral leukoplakia, a premalignant lesion. *J. Natl. Cancer Inst.* 85: 1951-1954.
37. LEE, J.J., W.K. HONG, W.N., HITTELMAN, *et al.* 2000. Predicting cancer development in oral leukoplakia: ten years of translational research. *Clin. Cancer Res.* 6: 1702-1710.
38. HITTELMAN W.N., R. YU, J. KURIE, *et al.* 1997. Evidence for genomic instability and clonal outgrowth in the bronchial epithelium of smokers [abstract]. *Proc. Am. Assoc. Cancer Res.* 38: 3097.
39. KURIE, J.M., J.S. LEE, F.R. KHURI, *et al.* N-(4-hydroxyphenyl)retinamide in the chemoprevention of squamous metaplasia and dysplasia of the bronchial epithelium. 2000. *Clin. Cancer Res.* 6: 2973-2979.
40. HITTELMAN, W.N., J.S. LEE, R.C. MORICE, *et al.* 1999. Lack of biomarker modulation in bronchial biopsies of chronic smokers following treatment with N-(4-hydroxyphenyl)retinamide (4-HPR). *Proc. Am. Assoc. Cancer Res.* 40: 2837.
41. HITTELMAN, W.N., J.S. LEE, N. CHEONG, *et al.* 1991. The chromosome view of "field cancerization" and multistep carcinogenesis. Implications for chemopreventive approaches. *In Chemoprevention of Cancer.* V. Pastorino & W.K. Hong, Eds.: 41-47. Georg Thieme Verlag. Stuttgart, Germany.
42. HITTELMAN, W.N., J.J. LEE, J.S. LEE, *et al.* 1998. Persistent genetic instability despite decreased proliferation in human lung tissue following smoking cessation. *Proc. AACR* 39: 336.

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Ashkenazi et al.	Group Art Unit: 1647
Serial No.: 09/903,925	Examiner: Fozia Hamid
Filed: July 11, 2001	<b>CERTIFICATE OF MAILING</b> I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Assistant Commissioner of Patents, Washington, D.C. 20231 on _____ Date: _____
For: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS	

DECLARATION OF AUDREY D. GODDARD, Ph.D UNDER 37 C.F.R. § 1.132

Assistant Commissioner of Patents  
Washington, D.C. 20231

Sir:

I, Audrey D. Goddard, Ph.D. do hereby declare and say as follows:

1. I am a Senior Clinical Scientist at the Experimental Medicine/BioOncology, Medical Affairs Department of Genentech, Inc., South San Francisco, California 94080.
2. Between 1993 and 2001, I headed the DNA Sequencing Laboratory at the Molecular Biology Department of Genentech, Inc. During this time, my responsibilities included the identification and characterization of genes contributing to the oncogenic process, and determination of the chromosomal localization of novel genes.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

Serial No.: \*

Filed: \*

4. I am familiar with a variety of techniques known in the art for detecting and quantifying the amplification of oncogenes in cancer, including the quantitative TaqMan PCR (i.e., "gene amplification") assay described in the above captioned patent application.

5. The TaqMan PCR assay is described, for example, in the following scientific publications: Higuchi *et al.*, Biotechnology 10:413-417 (1992) (Exhibit B); Livak *et al.*, PCR Methods Appl. 4:357-362 (1995) (Exhibit C) and Heid *et al.*, Genome Res. 6:986-994 (1996) (Exhibit D). Briefly, the assay is based on the principle that successful PCR yields a fluorescent signal due to Taq DNA polymerase-mediated exonuclease digestion of a fluorescently labeled oligonucleotide that is homologous to a sequence between two PCR primers. The extent of digestion depends directly on the amount of PCR, and can be quantified accurately by measuring the increment in fluorescence that results from decreased energy transfer. This is an extremely sensitive technique, which allows detection in the exponential phase of the PCR reaction and, as a result, leads to accurate determination of gene copy number.

6. The quantitative fluorescent TaqMan PCR assay has been extensively and successfully used to characterize genes involved in cancer development and progression. Amplification of protooncogenes has been studied in a variety of human tumors, and is widely considered as having etiological, diagnostic and prognostic significance. This use of the quantitative TaqMan PCR assay is exemplified by the following scientific publications: Pennica *et al.*, Proc. Natl. Acad. Sci. USA 95(25):14717-14722 (1998) (Exhibit E); Pitti *et al.*, Nature 396(6712):699-703 (1998) (Exhibit F) and Bieche *et al.*, Int. J. Cancer 78:661-666 (1998) (Exhibit G), the first two of which I am co-author. In particular, Pennica *et al.* have used the quantitative TaqMan PCR assay to study relative gene amplification of WISP and c-myc in various cell lines, colorectal tumors and normal mucosa. Pitti *et al.* studied the genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer, using the quantitative TaqMan PCR assay. Bieche *et al.* used the assay to study gene amplification in breast cancer.

Serial No.: \*

Filed: \*

7. It is my personal experience that the quantitative TaqMan PCR technique is technically sensitive enough to detect at least a 2-fold increase in gene copy number relative to control. It is further my considered scientific opinion that an at least 2-fold increase in gene copy number in a tumor tissue sample relative to a normal (i.e., non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

8. I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I declare that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Jan. 16, 2003

Date

Audrey D. Goddard

Audrey D. Goddard, Ph.D.

**AUDREY D. GODDARD, Ph.D.**

Genentech, Inc.  
1 DNA Way  
South San Francisco, CA, 94080  
650.225.6429  
goddarda@gene.com

110 Congo St.  
San Francisco, CA, 94131  
415.841.9154  
415.819.2247 (mobile)  
agoddard@pacbell.net

**PROFESSIONAL EXPERIENCE**

**Genentech, Inc.**  
**South San Francisco, CA**

**1993-present**

**2001 - present      Senior Clinical Scientist**  
Experimental Medicine / BioOncology, Medical Affairs

**Responsibilities:**

- *Companion diagnostic oncology products*
- *Acquisition of clinical samples from Genentech's clinical trials for translational research*
- *Translational research using clinical specimen and data for drug development and diagnostics*
- *Member of Development Science Review Committee, Diagnostic Oversight Team, 21 CFR Part 11 Subteam*

**Interests:**

- *Ethical and legal implications of experiments with clinical specimens and data*
- *Application of pharmacogenomics in clinical trials*

**1998 - 2001      Senior Scientist**  
Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

**Responsibilities:**

- *Management of a laboratory of up to nineteen –including postdoctoral fellow, associate scientist, senior research associate and research assistants/associate levels*
- *Management of a \$750K budget*
- *DNA sequencing core facility supporting a 350+ person research facility.*
- *DNA sequencing for high throughput gene discovery, - ESTs, cDNAs, and constructs*
- *Genomic sequence analysis and gene identification*
- *DNA sequence and primary protein analysis*

**Research:**

- *Chromosomal localization of novel genes*
- *Identification and characterization of genes contributing to the oncogenic process*
- *Identification and characterization of genes contributing to inflammatory diseases*
- *Design and development of schemes for high throughput genomic DNA sequence analysis*
- *Candidate gene prediction and evaluation*

**1993 - 1998      Scientist**

Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

**Responsibilities**

- *DNA sequencing core facility supporting a 350+ person research facility*
- *Assumed responsibility for a pre-existing team of five technicians and expanded the group into fifteen, introducing a level of middle management and additional areas of research*
- *Participated in the development of the basic plan for high throughput secreted protein discovery program – sequencing strategies, data analysis and tracking, database design*
- *High throughput EST and cDNA sequencing for new gene identification.*
- *Design and implementation of analysis tools required for high throughput gene identification.*
- *Chromosomal localization of genes encoding novel secreted proteins.*

**Research:**

- *Genomic sequence scanning for new gene discovery.*
- *Development of signal peptide selection methods.*
- *Evaluation of candidate disease genes.*
- *Growth hormone receptor gene SNPs in children with Idiopathic short stature*

**Imperial Cancer Research Fund  
London, UK with Dr. Ellen Solomon**

**1989-1992**

**6/89 – 12/92 Postdoctoral Fellow**

- Cloning and characterization of the genes fused at the acute promyelocytic leukemia translocation breakpoints on chromosomes 17 and 15.
- Prepared a successfully funded European Union multi-center grant application

**McMaster University  
Hamilton, Ontario, Canada with Dr. G. D. Sweeney**

**1983**

**5/83 – 8/83: NSERC Summer Student**

- *In vitro* metabolism of  $\beta$ -naphthoflavone in C57BL/6J and DBA mice

**EDUCATION**

**Ph.D.**

"Phenotypic and genotypic effects of mutations in the human retinoblastoma gene."

**Supervisor:** Dr. R. A. Phillips

University of Toronto  
Toronto, Ontario, Canada.  
Department of Medical  
Biophysics.

1989

**Honours B.Sc**

"The *in vitro* metabolism of the cytochrome P-448 inducer  $\beta$ -naphthoflavone in C57BL/6J mice."

**Supervisor:** Dr. G. D. Sweeney

McMaster University,  
Hamilton, Ontario, Canada.  
Department of Biochemistry

1983



## ACADEMIC AWARDS

Imperial Cancer Research Fund Postdoctoral Fellowship	1989-1992
Medical Research Council Studentship	1983-1988
NSERC Undergraduate Summer Research Award	1983
Society of Chemical Industry Merit Award (Hons. Biochem.)	1983
Dr. Harry Lyman Hooker Scholarship	1981-1983
J.L.W. Gill Scholarship	1981-1982
Business and Professional Women's Club Scholarship	1980-1981
Wyerhauser Foundation Scholarship	1979-1980

## INVITED PRESENTATIONS

Genentech's gene discovery pipeline: High throughput identification, cloning and characterization of novel genes. Functional Genomics: From Genome to Function, Litchfield Park, AZ, USA. October 2000

High throughput identification, cloning and characterization of novel genes. G2K:Back to Science, Advances in Genome Biology and Technology I. Marco Island, FL, USA. February 2000

Quality control in DNA Sequencing: The use of Phred and Phrap. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. April 1999

High throughput secreted protein identification and cloning. Tenth International Genome Sequencing and Analysis Conference, Miami, FL, USA. September 1998

The evolution of DNA sequencing: The Genentech perspective. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. May 1998

Partial Growth Hormone Insensitivity: The role of GH-receptor mutations in Idiopathic Short Stature. Tenth Annual National Cooperative Growth Study Investigators Meeting, San Francisco, CA, USA. October, 1996

Growth hormone (GH) receptor defects are present in selected children with non-GH-deficient short stature: A molecular basis for partial GH-insensitivity. 76<sup>th</sup> Annual Meeting of The Endocrine Society, Anaheim, CA, USA. June 1994

A previously uncharacterized gene, myl, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. XV International Association for Comparative Research on Leukemia and Related Disease, Padua, Italy. October 1991

## PATENTS

Goddard A, Godowski PJ, Gurney AL. NL2 Tie ligand homologue polypeptide. Patent Number: 6,455,496. Date of Patent: Sept. 24, 2002.

**Goddard A**, Godowski PJ and Gurney AL. NL3 Tie ligand homologue nucleic acids. Patent Number: 6,426,218. Date of Patent: July 30, 2002.

Godowski P, Gurney A, Hillan KJ, Botstein D, **Goddard A**, Roy M, Ferrara N, Tumas D, Schwall R. NL4 Tie ligand homologue nucleic acid. Patent Number: 6,4137,770. Date of Patent: July 2, 2002.

Ashkenazi A, Fong S, **Goddard A**, Gurney AL, Napier MA, Tumas D, Wood WI. Nucleic acid encoding A-33 related antigen poly peptides. Patent Number: 6,410,708. Date of Patent: Jun. 25, 2002.

Botstein DA, Cohen RL, **Goddard AD**, Gurney AL, Hillan KJ, Lawrence DA, Levine AJ, Pennica D, Roy MA and Wood WI. WISP polypeptides and nucleic acids encoding same. Patent Number: 6,387,657. Date of Patent: May 14, 2002.

**Goddard A**, Godowski PJ and Gurney AL. Tie ligands. Patent Number: 6,372,491. Date of Patent: April 16, 2002.

Godowski PJ, Gurney AL, **Goddard A** and Hillan K. TIE ligand homologue antibody. Patent Number: 6,350,450. Date of Patent: Feb. 26, 2002.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Tie receptor tyrosine kinase ligand homologues. Patent Number: 6,348,351. Date of Patent: Feb. 19, 2002.

**Goddard A**, Godowski PJ and Gurney AL. Ligand homologues. Patent Number: 6,348,350. Date of Patent: Feb. 19, 2002.

Attie KM, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 6,207,640. Date of Patent: March 27, 2001.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Nucleic acids encoding NL-3. Patent Number: 6,074,873. Date of Patent: June 13, 2000

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,824,642. Date of Patent: October 20, 1998

Attie K, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,646,113. Date of Patent: July 8, 1997

Multiple additional provisional applications filed

## PUBLICATIONS

Seshasayee D, Dowd P, Gu Q, Erickson S, **Goddard AD** Comparative sequence analysis of the *HER2* locus in mouse and man. Manuscript in preparation.

Abuzzahab MJ, **Goddard A**, Grigorescu F, Lautier C, Smith RJ and Chernausek SD. Human IGF-1 receptor mutations resulting in pre- and post-natal growth retardation. Manuscript in preparation.

Aggarwal S, Xie, M-H, Foster J, Frantz G, Stinson J, Corpuz RT, Simmons L, Hillan K, Yansura DG, Vandlen RL, **Goddard AD** and Gurney AL. FHFR, a novel receptor for the fibroblast growth factors. Manuscript submitted.

Adams SH, Chui C, Schilbach SL, Yu XX, **Goddard AD**, Grimaldi JC, Lee J, Dowd P, Colman S., Lewin DA. (2001) BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: Cloning, organization of the human gene, and assessment of a potential link to obesity. *Biochemical Journal* **360**: 135-142.

Lee J, Ho WH, Maruoka M, Corpuz RT, Baldwin DT, Foster JS, **Goddard AD**, Yansura DG, Vandlen RL, Wood WI, Gurney AL. (2001) IL-17E, a novel proinflammatory ligand for the IL-17 receptor homolog IL-17Rh1. *Journal of Biological Chemistry* **276**(2): 1660-1664.

Xie M-H, Aggarwal S, Ho W-H, Foster J, Zhang Z, Stinson J, Wood WI, **Goddard AD** and Gurney AL. (2000) Interleukin (IL)-22, a novel human cytokine that signals through the interferon-receptor related proteins CRF2-4 and IL-22R. *Journal of Biological Chemistry* **275**: 31335-31339.

Weiss GA, Watanabe CK, Zhong A, **Goddard A** and Sidhu SS. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA* **97**: 8950-8954.

Guo S, Yamaguchi Y, Schilbach S, Wada T., Lee J, **Goddard A**, French D, Handa H, Rosenthal A. (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**: 366-369.

Yan M, Wang L-C, Hymowitz SG, Schilbach S, Lee J, **Goddard A**, de Vos AM, Gao WQ, Dixit VM. (2000) Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**: 523-527.

Sehl PD, Tai JTN, Hillan KJ, Brown LA, **Goddard A**, Yang R, Jin H and Lowe DG. (2000) Application of cDNA microarrays in determining molecular phenotype in cardiac growth, development, and response to injury. *Circulation* **101**: 1990-1999.

Guo S, Brush J, Teraoka H, **Goddard A**, Wilson SW, Mullins MC and Rosenthal A. (1999) Development of noradrenergic neurons in the zebrafish hindbrain requires BMP, FGF8, and the homeodomain protein soulless/Phox2A. *Neuron* **24**: 555-566.

Stone D, Murone, M, Luoh, S, Ye W, Armanini P, Gurney A, Phillips HS, Brush, J, **Goddard A**, de Sauvage FJ and Rosenthal A. (1999) Characterization of the human suppressor of fused; a negative regulator of the zinc-finger transcription factor Gli. *J. Cell Sci.* **112**: 4437-4448.

Xie M-H, Holcomb I, Deuel B, Dowd P, Huang A, Vagts A, Foster J, Liang J, Brush J, Gu Q, Hillan K, **Goddard A** and Gurney, A.L. (1999) FGF-19, a novel fibroblast growth factor with unique specificity for FGFR4. *Cytokine* **11**: 729-735.

- Yan M, Lee J, Schilbach S, **Goddard A** and Dixit V. (1999) mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J. Biol. Chem.* **274**(15): 10287-10292.
- Gurney AL, Marsters SA, Huang RM, Pitti RM, Mark DT, Baldwin DT, Gray AM, Dowd P, Brush J, Heldens S, Schow P, **Goddard AD**, Wood WI, Baker KP, Godowski PJ and Ashkenazi A. (1999) Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Current Biology* **9**(4): 215-218.
- Ridgway JBB, Ng E, Kern JA, Lee J, Brush J, **Goddard A** and Carter P. (1999) Identification of a human anti-CD55 single-chain Fv by subtractive panning of a phage library using tumor and nontumor cell lines. *Cancer Research* **59**: 2718-2723.
- Pitti RM, Marsters SA, Lawrence DA, Roy M, Kischkel FC, Dowd P, Huang A, Donahue CJ, Sherwood SW, Baldwin DT, Godowski PJ, Wood WI, Gurney AL, Hillan KJ, Cohen RL, **Goddard AD**, Botstein D and Ashkenazi A. (1998) Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**(6712): 699-703.
- Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, Brush J, Taneyhill LA, Deuel B, Lew M, Watanabe C, Cohen RL, Melhem MF, Finley GG, Quirke P, **Goddard AD**, Hillan KJ, Gurney AL, Botstein D and Levine AJ. (1998) WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. *Proc. Natl. Acad. Sci. USA.* **95**(25): 14717-14722.
- Yang RB, Mark MR, Gray A, Huang A, Xie MH, Zhang M, **Goddard A**, Wood WI, Gurney AL and Godowski PJ. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* **395**(6699): 284-288.
- Merchant AM, Zhu Z, Yuan JQ, **Goddard A**, Adams CW, Presta LG and Carter P. (1998) An efficient route to human bispecific IgG. *Nature Biotechnology* **16**(7): 677-681.
- Marsters SA, Sheridan JP, Pitti RM, Brush J, **Goddard A** and Ashkenazi A. (1998) Identification of a ligand for the death-domain-containing receptor Apo3. *Current Biology* **8**(9): 525-528.
- Xie J, Murone M, Luoh SM, Ryan A, Gu Q, Zhang C, Bonifas JM, Lam CW, Hynes M, **Goddard A**, Rosenthal A, Epstein EH Jr. and de Sauvage FJ. (1998) Activating Smoothed mutations in sporadic basal-cell carcinoma. *Nature.* **391**(6662): 90-92.
- Marsters SA, Sheridan JP, Pitti RM, Huang A, Skubatch M, Baldwin D, Yuan J, Gurney A, **Goddard AD**, Godowski P and Ashkenazi A. (1997) A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Current Biology.* **7**(12): 1003-1006.
- Hynes M, Stone DM, Dowd M, Pitts-Meek S, **Goddard A**, Gurney A and Rosenthal A. (1997) Control of cell pattern in the neural tube by the zinc finger transcription factor *Gli-1*. *Neuron* **19**: 15-26.
- Sheridan JP, Marsters SA, Pitti RM, Gurney A., Skubatch M, Baldwin D, Ramakrishnan L, Gray CL, Baker K, Wood WI, **Goddard AD**, Godowski P, and Ashkenazi A. (1997) Control of TRAIL-Induced Apoptosis by a Family of Signaling and Decoy Receptors. *Science* **277** (5327): 818-821.

**Goddard AD**, Dowd P, Chernauek S, Geffner M, Gertner J, Hintz R, Hopwood N, Kaplan S, Plotnick L, Rogol A, Rosenfield R, Saenger P, Mauras N, Hershkopf R, Angulo M and Attie, K. (1997) Partial growth hormone insensitivity: The role of growth hormone receptor mutations in idiopathic short stature. *J. Pediatr.* **131**: S51-55.

Klein RD, Sherman D, Ho WH, Stone D, Bennett GL, Moffat B, Vandlen R, Simmons L, Gu Q, Hongo JA, Devaux B, Poulsen K, Armanini M, Nozaki C, Asai N, **Goddard A**, Phillips H, Henderson CE, Takahashi M and Rosenthal A. (1997) A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor. *Nature*. **387**(6634): 717-21.

Stone DM, Hynes M, Armanini M, Swanson TA, Gu Q, Johnson RL, Scott MP, Pennica D, **Goddard A**, Phillips H, Noll M, Hooper JE, de Sauvage F and Rosenthal A. (1996) The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**(6605): 129-34.

Marsters SA, Sheridan JP, Donahue CJ, Pitti RM, Gray CL, **Goddard AD**, Bauer KD and Ashkenazi A. (1996) Apo-3, a new member of the tumor necrosis factor receptor family, contains a death domain and activates apoptosis and NF-kappa  $\beta$ . *Current Biology* **6**(12): 1669-76.

Rothe M, Xiong J, Shu HB, Williamson K, **Goddard A** and Goeddel DV. (1996) I-TRAF is a novel TRAF-interacting protein that regulates TRAF-mediated signal transduction. *Proc. Natl. Acad. Sci. USA* **93**: 8241-8246.

Yang M, Luoh SM, **Goddard A**, Reilly D, Henzel W and Bass S. (1996) The bglX gene located at 47.8 min on the Escherichia coli chromosome encodes a periplasmic beta-glucosidase. *Microbiology* **142**: 1659-65.

**Goddard AD** and Black DM. (1996) Familial Cancer in Molecular Endocrinology of Cancer. Waxman, J. Ed. Cambridge University Press, Cambridge UK, pp.187-215.

Treanor JJS, Goodman L, de Sauvage F, Stone DM, Poulson KT, Beck CD, Gray C, Armanini MP, Pollocks RA, Hefti F, Phillips HS, **Goddard A**, Moore MW, Buj-Bello A, Davis AM, Asai N, Takahashi M, Vandlen R, Henderson CE and Rosenthal A. (1996) Characterization of a receptor for GDNF. *Nature* **382**: 80-83.

Klein RD, Gu Q, **Goddard A** and Rosenthal A. (1996) Selection for genes encoding secreted proteins and receptors. *Proc. Natl. Acad. Sci. USA* **93**: 7108-7113.

Winslow JW, Moran P, Valverde J, Shih A, Yuan JQ, Wong SC, Tsai SP, **Goddard A**, Henzel WJ, Hefti F and Caras I. (1995) Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved in axon bundle formation. *Neuron* **14**: 973-981.

Bennett BD, Zeigler FC, Gu Q, Fendly B, **Goddard AD**, Gillett N and Matthews W. (1995) Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kinase Htk. *Proc. Natl. Acad. Sci. USA* **92**: 1866-1870.

Huang X, Yuang J, **Goddard A**, Foulis A, James RF, Lernmark A, Pujol-Borrell R, Rabinovitch A, Somoza N and Stewart TA. (1995) Interferon expression in the pancreases of patients with type I diabetes. *Diabetes* **44**: 658-664.

**Goddard AD**, Yuan JQ, Fairbairn L, Dexter M, Borrow J, Kozak C and Solomon E. (1995) Cloning of the murine homolog of the leukemia-associated PML gene. *Mammalian Genome* **6**: 732-737.

- Goddard AD, Covello R, Luoh SM, Clackson T, Attie KM, Gesundheit N, Rundle AC, Wells JA, Carlsson LMTI and The Growth Hormone Insensitivity Study Group. (1995) Mutations of the growth hormone receptor in children with idiopathic short stature. *N. Engl. J. Med.* **333**: 1093-1098.
- Kuo SS, Moran P, Gripp J, Armanini M, Phillips HS, **Goddard A** and Caras IW. (1994) Identification and characterization of Batk, a predominantly brain-specific non-receptor protein tyrosine kinase related to Csk. *J. Neurosci. Res.* **38**: 705-715.
- Mark MR, Scadden DT, Wang Z, Gu Q, **Goddard A** and Godowski PJ. (1994) Rse, a novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *Journal of Biological Chemistry* **269**: 10720-10728.
- Borrow J, Shipley J, Howe K, Kiely F, **Goddard A**, Sheer D, Srivastava A, Antony AC, Fioretos T, Mitelman F and Solomon E. (1994) Molecular analysis of simple variant translocations in acute promyelocytic leukemia. *Genes Chromosomes Cancer* **9**: 234-243.
- Goddard AD** and Solomon E. (1993) Genetics of Cancer. *Adv. Hum. Genet.* **21**: 321-376.
- Borrow J, **Goddard AD**, Gibbons B, Katz F, Swirsky D, Fioretos T, Dube I, Winfield DA, Kingston J, Hagemeijer A, Rees JKH, Lister AT and Solomon E. (1992) Diagnosis of acute promyelocytic leukemia by RT-PCR: Detection of *PML-RARA* and *RARA-PML* fusion transcripts. *Br. J. Haematol.* **82**: 529-540.
- Goddard AD**, Borrow J and Solomon E. (1992) A previously uncharacterized gene, PML, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. *Leukemia* **6 Suppl 3**: 117S-119S.
- Zhu X, Dunn JM, **Goddard AD**, Squire JA, Becker A, Phillips RA and Gallie BL. (1992) Mechanisms of loss of heterozygosity in retinoblastoma. *Cytogenet. Cell. Genet.* **59**: 248-252.
- Foulkes W, **Goddard A.** and Patel K. (1991) Retinoblastoma linked with Seascale [letter]. *British Med. J.* **302**: 409.
- Goddard AD**, Borrow J, Freemont PS and Solomon E. (1991) Characterization of a novel zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science* **254**: 1371-1374.
- Solomon E, Borrow J and **Goddard AD**. (1991) Chromosomal aberrations in cancer. *Science* **254**: 1153-1160.
- Pajunen L, Jones TA, **Goddard A**, Sheer D, Solomon E, Pihlajaniemi T and Kivirikko KI. (1991) Regional assignment of the human gene coding for a multifunctional peptide (P4HB) acting as the  $\beta$ -subunit of prolyl-4-hydroxylase and the enzyme protein disulfide isomerase to 17q25. *Cytogenet. Cell. Genet.* **56**: 165-168.
- Borrow J, Black DM, **Goddard AD**, Yagle MK, Frischauf A.-M and Solomon E. (1991) Construction and regional localization of a *NotI* linking library from human chromosome 17q. *Genomics* **10**: 477-480.
- Borrow J, **Goddard AD**, Sheer D and Solomon E. (1990) Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. *Science* **249**: 1577-1580.

Myers JC, Jones TA, Pohjolainen E-R, Kadri AS, **Goddard AD**, Sheer D, Solomon E and Pihlajaniemi T. (1990) Molecular cloning of 5(IV) collagen and assignment of the gene to the region of the X-chromosome containing the Alport Syndrome locus. *Am. J. Hum. Genet.* **46**: 1024-1033.

Gallie BL, Squire JA, **Goddard A**, Dunn JM, Canton M, Hinton D, Zhu X and Phillips RA. (1990) Mechanisms of oncogenesis in retinoblastoma. *Lab. Invest.* **62**: 394-408.

**Goddard AD**, Phillips RA, Greger V, Passarge E, Hopping W, Gallie BL and Horsthemke B. (1990) Use of the RB1 cDNA as a diagnostic probe in retinoblastoma families. *Clinical Genetics* **37**: 117-126.

Zhu XP, Dunn JM, Phillips RA, **Goddard AD**, Paton KE, Becker A and Gallie BL. (1989) Germline, but not somatic, mutations of the RB1 gene preferentially involve the paternal allele. *Nature* **340**: 312-314.

Gallie BL, Dunn JM, **Goddard A**, Becker A and Phillips RA. (1988) Identification of mutations in the putative retinoblastoma gene. In Molecular Biology of The Eye: Genes, Vision and Ocular Disease. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 88. J. Piatigorsky, T. Shinohara and P.S. Zelenka, Eds. Alan R. Liss, Inc., New York, 1988, pp. 427-436.

**Goddard AD**, Balakier H, Canton M, Dunn J, Squire J, Reyes E, Becker A, Phillips RA and Gallie BL. (1988) Infrequent genomic rearrangement and normal expression of the putative RB1 gene in retinoblastoma tumors. *Mol. Cell. Biol.* **8**: 2082-2088.

Squire J, Dunn J, **Goddard A**, Hoffman T, Musarella M, Willard HF, Becker AJ, Gallie BL and Phillips RA. (1986) Cloning of the esterase D gene: A polymorphic gene probe closely linked to the retinoblastoma locus on chromosome 13. *Proc. Natl. Acad. Sci. USA* **83**: 6573-6577.

Squire J, **Goddard AD**, Canton M, Becker A, Phillips RA and Gallie BL (1986) Tumour induction by the retinoblastoma mutation is independent of N-myc expression. *Nature* **322**: 555-557.

**Goddard AD**, Heddle JA, Gallie BL and Phillips RA. (1985) Radiation sensitivity of fibroblasts of bilateral retinoblastoma patients as determined by micronucleus induction *in vitro*. *Mutation Research* **152**: 31-38.

otoxin fusion  
Sci. USA 85:

dmann, T. A.,  
immunotoxin  
to *Pseudomonas*

ngham, M. C.,  
od of cloning  
in *erb* as single-  
1066-1070.  
schberg, D. L.,  
our effects of  
topoisomerase phos-

s. G., Deleide,  
penetia, A. A.,  
re-targeted by  
1184-1189,  
text, Vol. 2, p.

and Stevens,  
with anti-viral  
J. Biochem. J.

A. I., Carnicelli,  
d properties of  
oxidase activity.

ization of the  
abius eukaryotic  
528.  
urification and  
stacea americana  
chem. Biophys.

1982. Purifica-  
of the antiviral  
'ekweed'. Bio-

L. Dodecaandrin,  
sodacandrin. Bio-

new inhibitor of  
tem. 255:6947-

Abbondanza, A.,  
ribosome-inacti-  
(white bryony).

-synthesis inhib-  
i Lett. 153:209-

8. Isolation and  
inhibitory pro-  
blem. 52:1229-

v. L., Sperti, S.,  
viro by procina  
elion). Biochem.

nza, A., Cenini,  
Purification and  
ith RNA N-gly-  
cation from the  
Acta. 993:287-

Guillemot, J. C.,  
1988. Trichoki-  
of *Trichostema*  
8.  
itors of animal  
Biophys. Acta

properties of the  
ocean inhibiting

of abrin: a toxic  
ferent biological  
ur. J. Biochem.

Freuz, M. 1980.  
i *Viscum album* L.

i. and Stirpe, F.,  
it modocan, the

., Brown, A. N.,  
s of volkenain, a  
14689-14995.  
nd properties of  
nary 18:2615-

# RESEARCH

## SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi\*, Gavin Dollinger<sup>1</sup>, P. Sean Walsh and Robert Griffith

Roche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. <sup>1</sup>Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. \*Corresponding author.

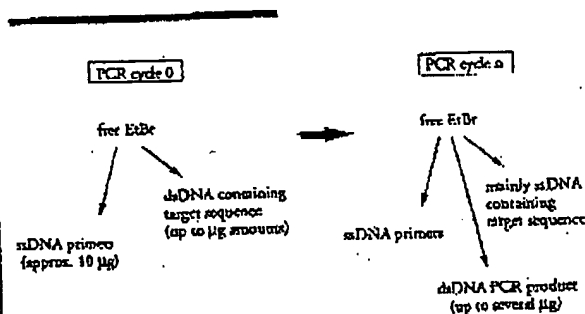
We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR<sup>1</sup> to clinical diagnostics are well known<sup>2,3</sup>, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases<sup>4</sup> made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization<sup>5,6</sup>, gel electrophoresis with or without use of restriction digestion<sup>7,8</sup>, HPLC<sup>9</sup>, or capillary electrophoresis<sup>10</sup>. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

"carryover" false positives in subsequent testing<sup>11</sup>.

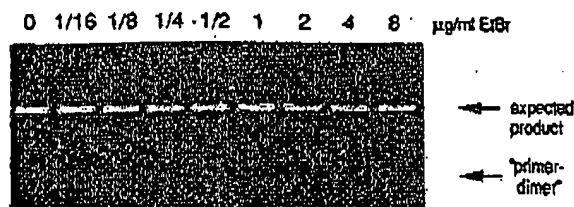
These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.<sup>12</sup>, developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.<sup>13</sup>, developed an assay in which the endogenous 5' exonuclease activity of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to ds-DNA<sup>14-16</sup>. As outlined in Figure 1, a prototypic PCR

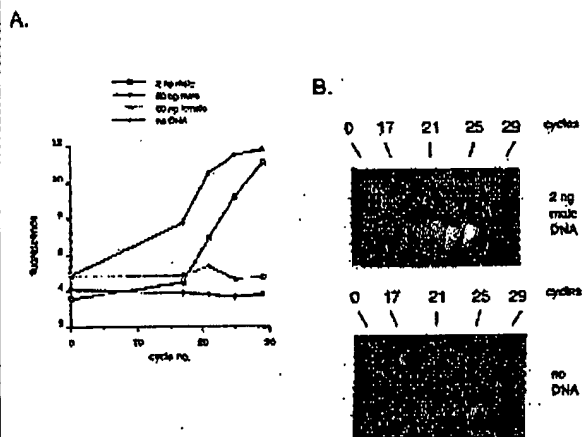


**FIGURE 1** Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.





**FIGURE 2** Gel electrophoresis of PCR amplification products of the human, nuclear gene, HLA DQ $\alpha$ , made in the presence of increasing amounts of EtBr (up to 8  $\mu$ g/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.



**FIGURE 3** (A) Fluorescence measurements from PCRs that contain 0.5  $\mu$ g/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photography of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ss-DNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA<sup>17</sup> to micrograms per PCR<sup>18</sup>. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

## RESULTS

**PCR in the presence of EtBr.** In order to assess the effect of EtBr in PCR, amplifications of the human HLA DQ $\alpha$  gene<sup>19</sup> were performed with the dye present at concentrations from 0.06 to 8.0  $\mu$ g/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5  $\mu$ g/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

**Detection of human Y-chromosome specific sequences.** Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5  $\mu$ g/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome<sup>20</sup>. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

**Detection of specific alleles of the human  $\beta$ -globin gene.** In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5  $\mu$ g/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human  $\beta$ -globin gene<sup>21</sup>. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the  $\beta$ -globin allele present<sup>22</sup>.

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type  $\beta$ -globin individual (AA); from a heterozygous sickle  $\beta$ -globin individual (AS); and from a homozygous sickle  $\beta$ -globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

emocy.

ess the  
HLA  
tent at  
oncen-  
lowing  
e 2, gel  
ie yield  
Br was  
ndicat-

se se-  
nent of  
ries of  
rimers  
human  
either  
DNA.  
fter 0,  
or each  
ts fluo-  
plotted  
of this  
case in  
DNA is  
umber.  
cc-fold  
ontain-  
ncrease  
her no  
DNA  
fewer  
in fluo-  
f these  
the ex-  
taining  
in the

ualized  
a UV  
b a red  
as that  
NA.  
-globin  
sch has  
etection  
Figure  
ications  
ography  
These  
for ci-  
human  
nparted  
ual 3'  
primer  
has am-  
c of the  
ent<sup>21,22</sup>  
nsist of  
fic (left  
i. Three  
zygous,  
ozygous  
ozygous  
genomic  
(3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a  $\beta$ -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both  $\beta$ -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for  $\beta$ -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

#### DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells<sup>6</sup>. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

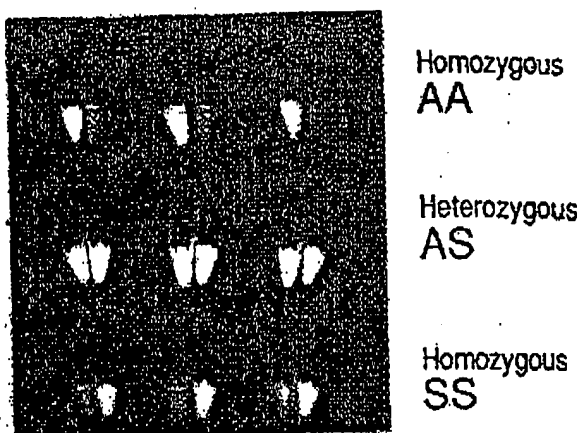


FIGURE 4 UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human  $\beta$ -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

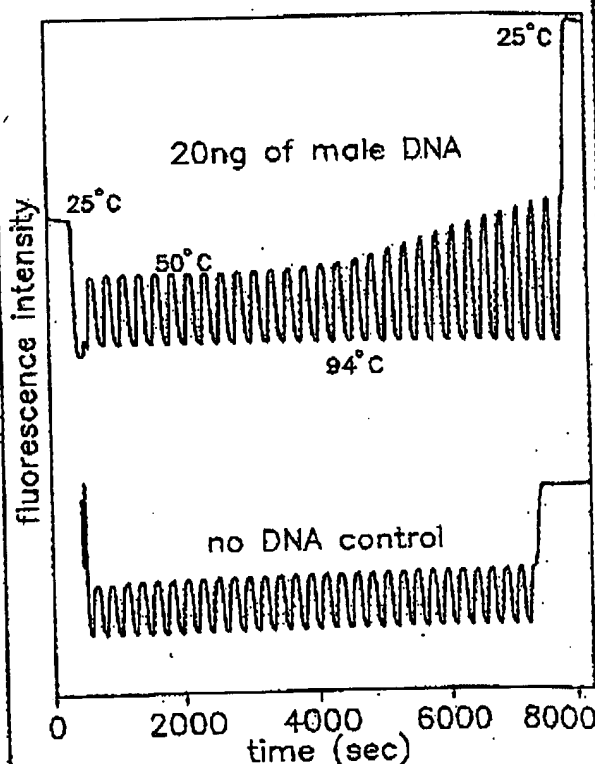


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube<sup>3</sup>, and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins<sup>23</sup>. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of  $10^5$  cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer<sup>24</sup>.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format<sup>25</sup>. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader<sup>26</sup>.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

#### EXPERIMENTAL PROTOCOL

**Human HLA-DQ $\alpha$  gene amplifications containing EtBr.** PCRs were set up in 100  $\mu$ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl<sub>2</sub>; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ $\alpha$  gene specific oligonucleotide primers (GH26 and GH27)<sup>19</sup> and approximately  $10^5$  copies of DQ $\alpha$  PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min, denaturation and 60°C for 30 sec, annealing and 72°C for 30 sec, extension.

**Y-chromosome specific PCR.** PCRs (100  $\mu$ l total reaction volume) containing 0.5  $\mu$ g/ml EtBr were prepared as described for HLA-DQ $\alpha$ , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2<sup>20</sup>, and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min, and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

**Allele-specific, human  $\beta$ -globin gene PCR.** Amplifications of 100  $\mu$ l volume using 0.5  $\mu$ g/ml of EtBr were prepared as described for HLA-DQ $\alpha$  above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HB14A (wild-type globin specific primers) or HGP2/HB14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.<sup>21</sup>. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min, and 55°C for 1 min, using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.<sup>21</sup> to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 23A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San Gabriel, CA).

**Fluorescence measurement.** Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPFX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Griest, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

**Continuous fluorescence monitoring of PCR.** Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPFX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

#### Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

#### References

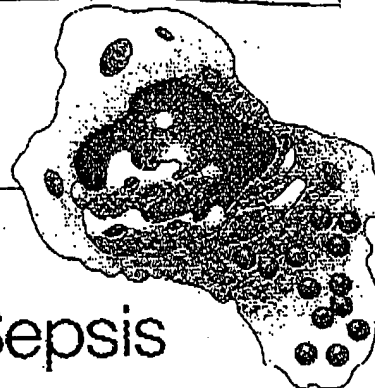
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHQB* 51:263-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Gelfand, D. and Smitsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1643-1651.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligomer cleavage detection. *J. Virol.* 61:1690-1694.
- Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:203-204.
- Horn, G. T., Richards, B. and Klugger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-555.
- Heiger, D. N., Cohen, A. S. and Rarger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-48.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. P. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Sier, W. 1979. Interactions of 4',6-diamidino-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:5519-5534.
- Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by <sup>1</sup>H NMR spectroscopy. *Nuc. Acids Res.* 18:3755-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 336:414-417.
- Abbott, M. A., Poiesz, B. J., Byrne, B. C., Kwok, S. Y., Smitsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified  $\beta$ -globin and HLA-DQA DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Kogan, S. C., Doherty, M. and Glusker, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:985-990.
- Wu, D. Y., Uguzzoni, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of  $\beta$ -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spacie, D., Gada, L., Levenson, C. and Smitsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Submitted.*
- Higuchi, R. 1989. Using PCR to engineer DNA. p. 61-70. In: *PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Pionta, E., Williams, J. F. and Woudenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-105, 106-112.
- Tumosa, N. and Kahan, L. 1989. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Meth.* 116:59-63.

# IBL

IMMUNO BIOLOGICAL LABORATORIES

## sCD-14 ELISA

### Trauma, Shock and Sepsis



The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations. During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features: 12x8 determinations (microtiter strips), precoated with a specific monoclonal antibody, 2x1 hour incubation, standard range: 3 - 96 ng/ml, detection limit: 1 ng/ml, CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNOCHEMIE UND -BIOLOGIE MBH  
OSTERSTRASSE 86 · D · 2000 HAMBURG 20 · GERMANY · TEL. +40/491 00 61-64 · FAX +40/40 11 98  
BIOTECHNOLOGY VOL 10 APRIL 1992

417

**GENENTECH, INC.**  
1 DNA Way  
South San Francisco, CA 94080 USA  
Phone: (650) 225-1000

---

FAX: (650) 952-9881

---

**FACSIMILE TRANSMITTAL**

---

**Date:** 19 July 2004

**To:** Anna Barry  
Heller Ehrman

**Re:** Higuchi reference

**Fax No:** 324-6638

**From:** Patty Tobin, Assistant to Elizabeth M. Barnes, Ph.D.  
Genentech, Inc. Legal Department

**Number of Pages including this cover sheet:** 6

## RESEARCH

## SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi\*, Gavin Dollinger<sup>1</sup>, P. Sean Walsh and Robert GriffithRoche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. <sup>1</sup>Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. \*Corresponding author.

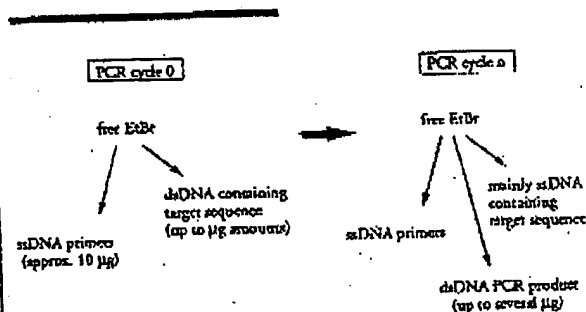
We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR<sup>1</sup> to clinical diagnostics are well known<sup>2,3</sup>, it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases<sup>4</sup> made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization<sup>5,6</sup>, gel electrophoresis with or without use of restriction digestion<sup>7,8</sup>, HPLC<sup>9</sup>, or capillary electrophoresis<sup>10</sup>. These methods are labor-intensive, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of

"carryover" false positives in subsequent testing<sup>11</sup>.

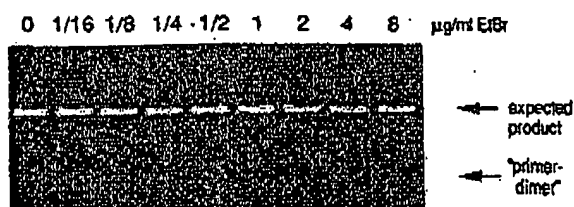
These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.<sup>12</sup>, developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.<sup>13</sup>, developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to dsDNA<sup>14-16</sup>. As outlined in Figure 1, a prototypic PCR

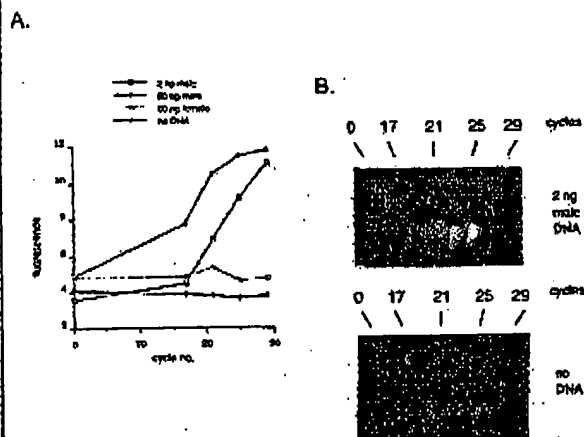


**FIGURE 1** Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.





**FIGURE 2** Gel electrophoresis of PCR amplification products of the human nuclear gene, HLA DQ $\alpha$ , made in the presence of increasing amounts of EtBr (up to 8  $\mu$ g/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.



**FIGURE 3** (A) Fluorescence measurements from PCRs that contain 0.5  $\mu$ g/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photograph of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ssDNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA<sup>17</sup> to micrograms per PCR<sup>18</sup>. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel

before and after, or even continuously during, thermocycling.

## RESULTS

**PCR in the presence of EtBr.** In order to assess the effect of EtBr in PCR, amplifications of the human HLA DQ $\alpha$  gene<sup>19</sup> were performed with the dye present at concentrations from 0.06 to 8.0  $\mu$ g/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5  $\mu$ g/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

**Detection of human Y-chromosome specific sequences.** Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5  $\mu$ g/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome<sup>20</sup>. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

**Detection of specific alleles of the human  $\beta$ -globin gene.** In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5  $\mu$ g/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human  $\beta$ -globin gene<sup>21</sup>. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the  $\beta$ -globin allele present<sup>21,22</sup>.

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous, wild-type  $\beta$ -globin individual (AA); from a heterozygous sickle  $\beta$ -globin individual (AS); and from a homozygous sickle  $\beta$ -globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

emocy.

ess the  
a HLA  
cent at  
oncen-  
lowing  
e 2, gel  
ic yield  
Br was  
ndicat-

5c se-  
nent of  
ries of  
rimers  
human  
either  
DNA.  
after 0,  
or each  
is fluo-  
plotted  
of this  
case in  
DNA is  
umber.  
re-fold  
ontain-  
crease  
her no  
DNA  
fewer  
in fluo-  
f these  
the ex-  
taining  
in the

ualized  
n a UV  
h a red  
ns that  
4A.

-globin  
sch has  
etection  
Figure  
ications  
ography  
These  
for ci-  
human  
nparted  
ional 3'  
primer  
has am-  
c of the  
ent<sup>21,22</sup>  
nsists of  
the (left  
Three  
zygous,  
ozygous  
ozygous  
genomic  
(3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a  $\beta$ -globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both  $\beta$ -globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for  $\beta$ -globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

## DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells<sup>6</sup>. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

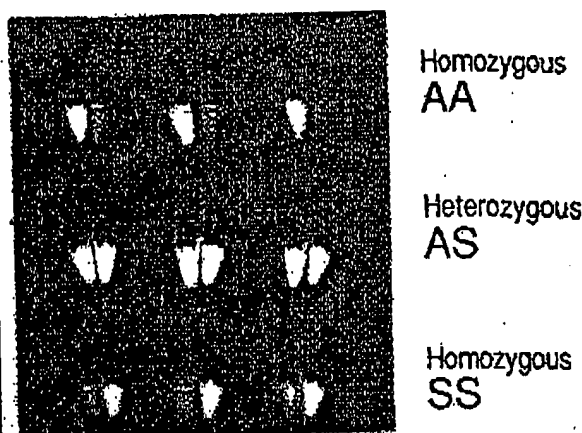


FIGURE 4 UV photograph of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human  $\beta$ -globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type alleles, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

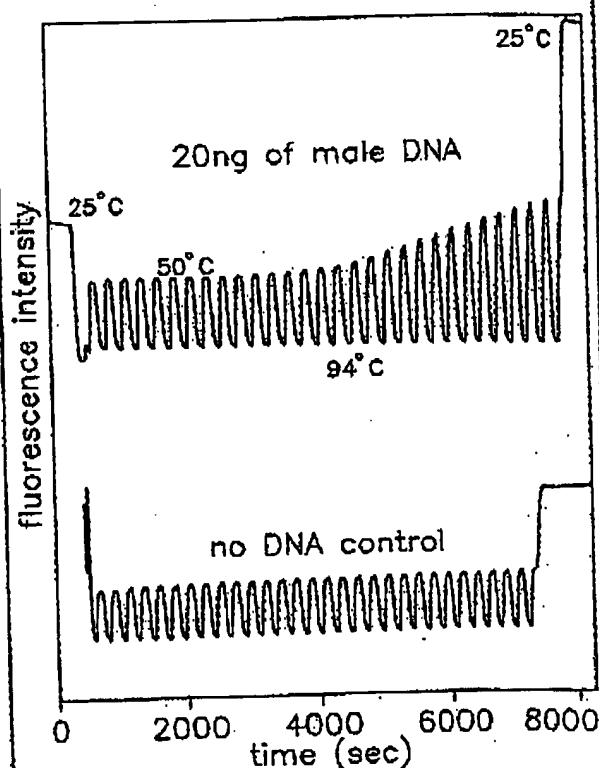


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.



DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube<sup>3</sup>, and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins<sup>23</sup>. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of  $10^5$  cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer<sup>24</sup>.

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format<sup>25</sup>. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader<sup>26</sup>.

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true

positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

#### EXPERIMENTAL PROTOCOL

**Human HLA-DQ $\alpha$  gene amplifications containing EtBr.** PCRs were set up in 100  $\mu$ l volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM MgCl<sub>2</sub>; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQ $\alpha$  gene specific oligonucleotide primers (GH26 and GH27<sup>19</sup> and approximately  $10^5$  copies of DQ $\alpha$  PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min, denaturation and 60°C for 30 sec, annealing and 72°C for 30 sec, extension.

**Y-chromosome specific PCR.** PCRs (100  $\mu$ l total reaction volume) containing 0.5  $\mu$ g/ml EtBr were prepared as described for HLA-DQ $\alpha$ , except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2<sup>20</sup>, and either 60 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min, and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

**Allele-specific, human  $\beta$ -globin gene PCR.** Amplifications of 100  $\mu$ l volume using 0.5  $\mu$ g/ml EtBr were prepared as described for HLA-DQ $\alpha$  above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HB14A (wild-type globin specific primers) or HGP2/HB14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.<sup>21</sup>. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min, and 55°C for 1 min, using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.<sup>21</sup> to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 23A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San Gabriel, CA).

**Fluorescence measurement.** Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Crist, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

**Continuous fluorescence monitoring of PCR.** Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

#### Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

#### References

- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *CSHSQB* 51:263-273.
- White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. *Trends Genet.* 5:185-189.
- Erlich, H. A., Gelfand, D. and Smitsky, J. J. 1991. Recent advances in the polymerase chain reaction. *Science* 252:1643-1651.
- Saiki, R. K., Gelfand, D. M., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.
- Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234.
- Kwok, S. Y., Mack, D. H., Mullis, K. B., Foksz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using *in vitro* enzymatic amplification and oligomer cleavage detection. *J. Virol.* 61:1690-1694.
- Chhab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemia. *Nature* 329:203-204.
- Horn, G. T., Richards, B. and Klinger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nuc. Acids Res.* 16:2140.
- Katz, E. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. *Biotechniques* 8:546-555.
- Heiger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. *J. Chromatogr.* 516:33-48.
- Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. *Nature* 339:237-238.
- Chhab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86:9178-9182.
- Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88:7276-7280.
- Markovits, J., Roques, B. F. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. *Anal. Biochem.* 94:259-264.
- Kapuscinski, J. and Sacer, W. 1979. Interactions of 4',6-diamidino-2-phenylindole with synthetic polynucleotides. *Nuc. Acids Res.* 6:5519-5534.
- Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by <sup>1</sup>H NMR spectroscopy. *Nuc. Acids Res.* 18:3753-3762.
- Li, H. H., Gyllenstein, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 336:414-417.
- Abbott, M. A., Potocz, B. J., Byrne, B. C., Kwok, S. Y., Smitsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified *in vitro*. *J. Infect. Dis.* 158:1158.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified  $\beta$ -globin and HLA-DQA DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
- Kogan, S. G., Doherty, M. and Giachini, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* 317:985-990.
- Wu, D. Y., Ugozzoli, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of  $\beta$ -globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* 86:2757-2760.
- Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Smitsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nuc. Acids Res.* 18:999-1005.
- Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Submitted.
- Higuchi, R. 1989. Using PCR to engineer DNA. p. 61-70. In: *PCR Technology*. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
- Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Fionna, E., Williams, J. F. and Woudenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. *Biotechniques* 10:102-105, 106-112.
- Tumosa, N. and Kalan, L. 1989. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. *J. Immun. Meth.* 116:59-63.

# IBL

IMMUNO BIOLOGICAL LABORATORIES

## sCD-14 ELISA

### Trauma, Shock and Sepsis

The CD-14 molecule is expressed on the surface of monocytes and some macrophages. Membrane-bound CD-14 is a receptor for lipopolysaccharide (LPS) complexed to LPS-Binding-Protein (LBP). The concentration of its soluble form is altered under certain pathological conditions. There is evidence for an important role of sCD-14 with polytrauma, sepsis, burnings and inflammations. During septic conditions and acute infections it seems to be a prognostic marker and is therefore of value in monitoring these patients.

IBL offers an ELISA for quantitative determination of soluble CD-14 in human serum, -plasma, cell-culture supernatants and other biological fluids.

Assay features:

- 12 x 8 determinations (microtiter strips),
- precoated with a specific monoclonal antibody,
- 2x1 hour incubation,
- standard range: 3 - 96 ng/ml
- detection limit: 1 ng/ml
- CV: intra- and interassay < 8%

For more information call or fax

GESELLSCHAFT FÜR IMMUNCHEMIE UND -BIOLOGIE MBH  
OSTERSTRASSE 86 · D-2000 HAMBURG 20 · GERMANY · TEL. +40/491 00 61-64 · FAX +40/40 11 98

BIO/TECHNOLOGY VOL 10 APRIL 1992

417

# Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization

Kenneth J. Livak, Susan J.A. Flood, Jeffrey Marmaro, William Giusti, and Karin Deetz

Perkin-Elmer, Applied Biosystems Division, Foster City, California 94404

The 5' nuclease PCR assay detects the accumulation of specific PCR product by hybridization and cleavage of a double-labeled fluorogenic probe during the amplification reaction. The probe is an oligonucleotide with both a reporter fluorescent dye and a quencher dye attached. An increase in reporter fluorescence intensity indicates that the probe has hybridized to the target PCR product and has been cleaved by the 5' → 3' nucleolytic activity of *Taq* DNA polymerase. In this study, probes with the quencher dye attached to an internal nucleotide were compared with probes with the quencher dye attached to the 3'-end nucleotide. In all cases, the reporter dye was attached to the 5' end. All intact probes showed quenching of the reporter fluorescence. In general, probes with the quencher dye attached to the 3'-end nucleotide exhibited a larger signal in the 5' nuclease PCR assay than the internally labeled probes. It is proposed that the larger signal is caused by increased likelihood of cleavage by *Taq* DNA polymerase when the probe is hybridized to a template strand during PCR. Probes with the quencher dye attached to the 3'-end nucleotide also exhibited an increase in reporter fluorescence intensity when hybridized to a complementary strand. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends can be used as homogeneous hybridization probes.

A homogeneous assay for detecting the accumulation of specific PCR product that uses a double-labeled fluorogenic probe was described by Lee et al.<sup>(1)</sup> The assay exploits the 5' → 3' nucleolytic activity of *Taq* DNA polymerase<sup>(2,3)</sup> and is diagrammed in Figure 1. The fluorogenic probe consists of an oligonucleotide with a reporter fluorescent dye, such as a fluorescein, attached to the 5' end; and a quencher dye, such as a rhodamine, attached internally. When the fluorescein is excited by irradiation, its fluorescent emission will be quenched if the rhodamine is close enough to be excited through the process of fluorescence energy transfer (FET).<sup>(4,5)</sup> During PCR, if the probe is hybridized to a template strand, *Taq* DNA polymerase will cleave the probe because of its inherent 5' → 3' nucleolytic activity. If the cleavage occurs between the fluorescein and rhodamine dyes, it causes an increase in fluorescein fluorescence intensity because the fluorescein is no longer quenched. The increase in fluorescein fluorescence intensity indicates that the probe-specific PCR product has been generated. Thus, FET between a reporter dye and a quencher dye is critical to the performance of the probe in the 5' nuclease PCR assay.

Quenching is completely dependent on the physical proximity of the two dyes.<sup>(6)</sup> Because of this, it has been assumed that the quencher dye must be attached near the 5' end. Surprisingly, we have found that attaching a rhodamine dye at the 3' end of a probe still provides adequate quenching for the probe to perform in the 5' nuclease

PCR assay. Furthermore, cleavage of this type of probe is not required to achieve some reduction in quenching. Oligonucleotides with a reporter dye on the 5' end and a quencher dye on the 3' end exhibit a much higher reporter fluorescence when double-stranded as compared with single-stranded. This should make it possible to use this type of double-labeled probe for homogeneous detection of nucleic acid hybridization.

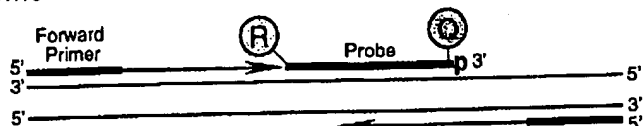
## MATERIALS AND METHODS

### Oligonucleotides

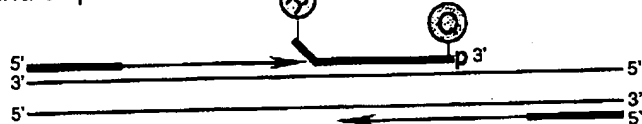
Table 1 shows the nucleotide sequence of the oligonucleotides used in this study. Linker arm nucleotide (LAN) phosphoramidite was obtained from Glen Research. The standard DNA phosphoramidites, 6-carboxyfluorescein (6-FAM) phosphoramidite, 6-carboxytetramethylrhodamine succinimidyl ester (TAMRA NHS ester), and Phosphalink for attaching a 3'-blocking phosphate, were obtained from Perkin-Elmer, Applied Biosystems Division. Oligonucleotide synthesis was performed using an ABI model 394 DNA synthesizer (Applied Biosystems). Primer and complement oligonucleotides were purified using Oligo Purification Cartridges (Applied Biosystems). Double-labeled probes were synthesized with 6-FAM-labeled phosphoramidite at the 5' end, LAN replacing one of the T's in the sequence, and Phosphalink at the 3' end. Following deprotection and ethanol precipitation, TAMRA NHS ester was coupled to the LAN-containing oligonucleotide in 250

# Research

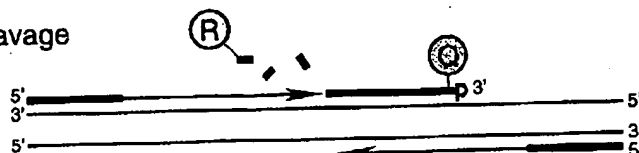
## Polymerization



## Strand displacement



## Cleavage



## Polymerization completed

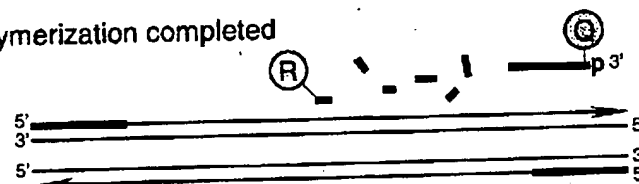


FIGURE 1 Diagram of 5' nuclease assay. Stepwise representation of the 5' → 3' nucleolytic activity of *Taq* DNA polymerase acting on a fluorogenic probe during one extension phase of PCR.

mm Na-bicarbonate buffer (pH 9.0) at room temperature. Unreacted dye was removed by passage over a PD-10 Sephadex column. Finally, the double-labeled probe was purified by preparative high-performance liquid chromatography (HPLC) using an Aquapore C<sub>8</sub> 220×4.6-mm column with 7-μm particle size. The column was developed with a 24-min linear gradient of 8–20% acetonitrile in 0.1 M TEAA (triethylamine acetate). Probes are named by designating the sequence from Table 1 and the position of the LAN-TAMRA moiety. For example, probe A1-7 has sequence A1 with LAN-TAMRA at nucleotide position 7 from the 5' end.

### PCR Systems

All PCR amplifications were performed in the Perkin-Elmer GeneAmp PCR System 9600 using 50-μl reactions that contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 400 μM dUTP, 0.5 unit of AmpErase uracil N-glycosylase (Perkin-Elmer), and 1.25 unit of AmpliTaq DNA polymerase (Perkin-Elmer). A 295-bp segment from exon 3 of the human β-actin

gene (nucleotides 2141–2435 in the sequence of Nakajima-Iijima et al.<sup>(7)</sup>) was amplified using primers AFP and ARP (Table 1), which are modified slightly from those of du Breuil et al.<sup>(8)</sup> Actin amplification reactions contained 4 mM MgCl<sub>2</sub>, 20 ng of human genomic DNA, 50 nM A1 or A3 probe, and 300 nM each

primer. The thermal regimen was 50°C (2 min), 95°C (10 min), 40 cycles of 95°C (20 sec), 60°C (1 min), and hold at 72°C. A 515-bp segment was amplified from a plasmid that consists of a segment of λ DNA (nucleotides 32,220–32,747) inserted in the *Sma*I site of vector pUC119. These reactions contained 3.5 mM MgCl<sub>2</sub>, 1 ng of plasmid DNA, 50 nM P2 or P5 probe, 200 nM primer F119, and 200 nM primer R119. The thermal regimen was 50°C (2 min), 95°C (10 min), 25 cycles of 95°C (20 sec), 57°C (1 min), and hold at 72°C.

### Fluorescence Detection

For each amplification reaction, a 40-μl aliquot of a sample was transferred to an individual well of a white, 96-well microtiter plate (Perkin-Elmer). Fluorescence was measured on the Perkin-Elmer TaqMan LS-50B System, which consists of a luminescence spectrometer with plate reader assembly, a 485-nm excitation filter, and a 515-nm emission filter. Excitation was at 488 nm using a 5-nm slit width. Emission was measured at 518 nm for 6-FAM (the reporter or R value) and 582 nm for TAMRA (the quencher or Q value) using a 10-nm slit width. To determine the increase in reporter emission that is caused by cleavage of the probe during PCR, three normalizations are applied to the raw emission data. First, emission intensity of a buffer blank is subtracted for each wavelength. Second, emission intensity of the reporter is

TABLE 1 Sequences of Oligonucleotides

Name	Type	Sequence
F119	primer	ACCCACAGGAAGTATGATCACCCTC
R119	primer	ATGTCGCGTTTCGGCTGACGTTCTGC
P2	probe	TCGCATTACTGATCGTTCGCCAACCAAGT <sub>p</sub>
P2C	complement	GTACTGGTTGGCAACGATCAGTAATGCGATC
P5	probe	CGGAATTGCTGGTATCTATGACAAGGAT <sub>p</sub>
P5C	complement	TTCATCCTTGTCTAGATACCAAGCAATCCG
AFP	primer	TCACCCACACTGTGCCCATCTACGA
ARP	primer	CAGCGGAACCGCTCATTGCCAATGG
A1	probe	ATGCCCTCCCCATGCCATCTCGG <sub>p</sub>
A1C	complement	AGACGCGAGGATGGCATGGGGGAGGGGCATAC
A3	probe	CGCCCTGGACTTCGAGCAAGAGAT <sub>p</sub>
A3C	complement	CCATCTCTTGCTCGAAGTCCAGGGCGAC

For each oligonucleotide used in this study, the nucleic acid sequence is given, written in the 5' → 3' direction. There are three types of oligonucleotides: PCR primer, fluorogenic probe used in the 5' nuclease assay, and complement used to hybridize to the corresponding probe. For the probes, the underlined base indicates a position where LAN with TAMRA attached was substituted for a T. (p) The presence of a 3' phosphate on each probe.

A1-2 RAQGGCCCTCCCCCATGCCATCCTGCGTp  
 A1-7 RATGCCCCQCCCCCATGCCATCCTGCGTp  
 A1-14 RATGCCCTCCCCCAQGCCATCCTGCGTp  
 A1-19 RATGCCCTCCCCCATGCCAQCTGCGTp  
 A1-22 RATGCCCTCCCCCATGCCATCCQCGTp  
 A1-26 RATGCCCTCCCCCATGCCATCCTGCGQp

Probe	518 nm		582 nm		RQ <sup>-</sup>	RQ <sup>+</sup>	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A1-2	25.5 ± 2.1	32.7 ± 1.9	38.2 ± 3.0	38.2 ± 2.0	0.67 ± 0.01	0.86 ± 0.06	0.19 ± 0.06
A1-7	53.5 ± 4.3	395.1 ± 21.4	108.5 ± 6.3	110.3 ± 5.3	0.49 ± 0.03	3.58 ± 0.17	3.08 ± 0.18
A1-14	127.0 ± 4.9	403.5 ± 19.1	108.7 ± 5.3	93.1 ± 6.3	1.16 ± 0.02	4.34 ± 0.15	3.18 ± 0.15
A1-19	187.5 ± 17.9	422.7 ± 7.7	70.3 ± 7.4	73.0 ± 2.8	2.67 ± 0.05	5.80 ± 0.15	3.13 ± 0.16
A1-22	224.6 ± 9.4	482.2 ± 43.6	100.0 ± 4.0	96.2 ± 9.6	2.25 ± 0.03	5.02 ± 0.11	2.77 ± 0.12
A1-26	160.2 ± 8.9	454.1 ± 18.4	93.1 ± 5.4	90.7 ± 3.2	1.72 ± 0.02	5.01 ± 0.08	3.29 ± 0.08

**FIGURE 2** Results of 5' nuclease assay comparing β-actin probes with TAMRA at different nucleotide positions. As described in Materials and Methods, PCR amplifications containing the indicated probes were performed, and the fluorescence emission was measured at 518 and 582 nm. Reported values are the average ± 1 S.D. for six reactions run without added template (no temp.) and six reactions run with template (+ temp.). The RQ ratio was calculated for each individual reaction and averaged to give the reported RQ<sup>-</sup> and RQ<sup>+</sup> values.

divided by the emission intensity of the quencher to give an RQ ratio for each reaction tube. This normalizes for well-to-well variations in probe concentration and fluorescence measurement. Finally, ΔRQ is calculated by subtracting the RQ value of the no-template control (RQ<sup>-</sup>) from the RQ value for the complete reaction including template (RQ<sup>+</sup>).

## RESULTS

A series of probes with increasing distances between the fluorescein reporter and rhodamine quencher were tested to investigate the minimum and maximum spacing that would give an acceptable performance in the 5' nuclease PCR assay. These probes hybridize to a target

sequence in the human β-actin gene. Figure 2 shows the results of an experiment in which these probes were included in PCR that amplified a segment of the β-actin gene containing the target sequence. Performance in the 5' nuclease PCR assay is monitored by the magnitude of ΔRQ, which is a measure of the increase in reporter fluorescence caused by PCR amplification of the probe target. Probe A1-2 has a ΔRQ value that is close to zero, indicating that the probe was not cleaved appreciably during the amplification reaction. This suggests that with the quencher dye on the second nucleotide from the 5' end, there is insufficient room for *Taq* polymerase to cleave efficiently between the reporter and quencher. The other five probes exhibited comparable ΔRQ values that are

clearly different from zero. Thus, all five probes are being cleaved during PCR amplification resulting in a similar increase in reporter fluorescence. It should be noted that complete digestion of a probe produces a much larger increase in reporter fluorescence than that observed in Figure 2 (data not shown). Thus, even in reactions where amplification occurs, the majority of probe molecules remain uncleaved. It is mainly for this reason that the fluorescence intensity of the quencher dye TAMRA changes little with amplification of the target. This is what allows us to use the 582-nm fluorescence reading as a normalization factor.

The magnitude of RQ<sup>-</sup> depends mainly on the quenching efficiency inherent in the specific structure of the probe and the purity of the oligonucleotide. Thus, the larger RQ<sup>-</sup> values indicate that probes A1-14, A1-19, A1-22, and A1-26 probably have reduced quenching as compared with A1-7. Still, the degree of quenching is sufficient to detect a highly significant increase in reporter fluorescence when each of these probes is cleaved during PCR.

To further investigate the ability of TAMRA on the 3' end to quench 6-FAM on the 5' end, three additional pairs of probes were tested in the 5' nuclease PCR assay. For each pair, one probe has TAMRA attached to an internal nucleotide and the other has TAMRA attached to the 3' end nucleotide. The results are shown in Table 2. For all three sets, the probe with the 3' quencher exhibits a ΔRQ value that is considerably higher than for the probe with the internal quencher. The RQ<sup>-</sup> values suggest that differences in quenching are not as great as those observed with some of the A1 probes. These results demonstrate that a quencher dye on the 3' end of an oligonucleotide can quench efficiently the

**TABLE 2** Results of 5' Nuclease Assay Comparing Probes with TAMRA Attached to an Internal or 3'-terminal Nucleotide

Probe	518 nm		582 nm		RQ <sup>-</sup>	RQ <sup>+</sup>	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A3-6	54.6 ± 3.2	84.8 ± 3.7	116.2 ± 6.4	115.6 ± 2.5	0.47 ± 0.02	0.73 ± 0.03	0.26 ± 0.04
A3-24	72.1 ± 2.9	236.5 ± 11.1	84.2 ± 4.0	90.2 ± 3.8	0.86 ± 0.02	2.62 ± 0.05	1.76 ± 0.05
P2-7	82.8 ± 4.4	384.0 ± 34.1	105.1 ± 6.4	120.4 ± 10.2	0.79 ± 0.02	3.19 ± 0.16	2.40 ± 0.16
P2-27	113.4 ± 6.6	555.4 ± 14.1	140.7 ± 8.5	118.7 ± 4.8	0.81 ± 0.01	4.68 ± 0.10	3.88 ± 0.10
P5-10	77.5 ± 6.5	244.4 ± 15.9	86.7 ± 4.3	95.8 ± 6.7	0.89 ± 0.05	2.55 ± 0.06	1.66 ± 0.08
P5-28	64.0 ± 5.2	333.6 ± 12.1	100.6 ± 6.1	94.7 ± 6.3	0.63 ± 0.02	3.53 ± 0.12	2.89 ± 0.13

Reactions containing the indicated probes and calculations were performed as described in Material and Methods and in the legend to Fig. 2.

## Research

fluorescence of a reporter dye on the 5' end. The degree of quenching is sufficient for this type of oligonucleotide to be used as a probe in the 5' nuclease PCR assay.

To test the hypothesis that quenching by a 3' TAMRA depends on the flexibility of the oligonucleotide, fluorescence was measured for probes in the single-stranded and double-stranded states. Table 3 reports the fluorescence observed at 518 and 582 nm. The relative degree of quenching is assessed by calculating the RQ ratio. For probes with TAMRA 6–10 nucleotides from the 5' end, there is little difference in the RQ values when comparing single-stranded with double-stranded oligonucleotides. The results for probes with TAMRA at the 3' end are much different. For these probes, hybridization to a complementary strand causes a dramatic increase in RQ. We propose that this loss of quenching is caused by the rigid structure of double-stranded DNA, which prevents the 5' and 3' ends from being in proximity.

When TAMRA is placed toward the 3' end, there is a marked  $Mg^{2+}$  effect on quenching. Figure 3 shows a plot of observed RQ values for the A1 series of probes as a function of  $Mg^{2+}$  concentration. With TAMRA attached near the 5' end (probe A1-2 or A1-7), the RQ value at 0 mM  $Mg^{2+}$  is only slightly higher than RQ at 10 mM  $Mg^{2+}$ . For probes A1-19, A1-22, and A1-26, the RQ values at 0 mM  $Mg^{2+}$  are very high, indicating a much

reduced quenching efficiency. For each of these probes, there is a marked decrease in RQ at 1 mM  $Mg^{2+}$  followed by a gradual decline as the  $Mg^{2+}$  concentration increases to 10 mM. Probe A1-14 shows an intermediate RQ value at 0 mM  $Mg^{2+}$  with a gradual decline at higher  $Mg^{2+}$  concentrations. In a low-salt environment with no  $Mg^{2+}$  present, a single-stranded oligonucleotide would be expected to adopt an extended conformation because of electrostatic repulsion. The binding of  $Mg^{2+}$  ions acts to shield the negative charge of the phosphate backbone so that the oligonucleotide can adopt conformations where the 3' end is close to the 5' end. Therefore, the observed  $Mg^{2+}$  effects support the notion that quenching of a 5' reporter dye by TAMRA at or near the 3' end depends on the flexibility of the oligonucleotide.

### DISCUSSION

The striking finding of this study is that it seems the rhodamine dye TAMRA, placed at any position in an oligonucleotide, can quench the fluorescent emission of a fluorescein (6-FAM) placed at the 5' end. This implies that a single-stranded, double-labeled oligonucleotide must be able to adopt conformations where the TAMRA is close to the 5' end. It should be noted that the decay of 6-FAM in the excited state requires a certain amount of time. Therefore, what

matters for quenching is not the average distance between 6-FAM and TAMRA but, rather, how close TAMRA can get to 6-FAM during the lifetime of the 6-FAM excited state. As long as the decay time of the excited state is relatively long compared with the molecular motions of the oligonucleotide, quenching can occur. Thus, we propose that TAMRA at the 3' end, or any other position, can quench 6-FAM at the 5' end because TAMRA is in proximity to 6-FAM often enough to be able to accept energy transfer from an excited 6-FAM.

Details of the fluorescence measurements remain puzzling. For example, Table 3 shows that hybridization of probes A1-26, A3-24, and P5-28 to their complementary strands not only causes a large increase in 6-FAM fluorescence at 518 nm but also causes a modest increase in TAMRA fluorescence at 582 nm. If TAMRA is being excited by energy transfer from quenched 6-FAM, then loss of quenching attributable to hybridization should cause a decrease in the fluorescence emission of TAMRA. The fact that the fluorescence emission of TAMRA increases indicates that the situation is more complex. For example, we have anecdotal evidence that the bases of the oligonucleotide, especially G, quench the fluorescence of both 6-FAM and TAMRA to some degree. When double-stranded, base-pairing may reduce the ability of the bases to quench. The primary factor causing the quenching of 6-FAM in an intact probe is the TAMRA dye. Evidence for the importance of TAMRA is that 6-FAM fluorescence remains relatively unchanged when probes labeled only with 6-FAM are used in the 5' nuclease PCR assay (data not shown). Secondary effectors of fluorescence, both before and after cleavage of the probe, need to be explored further.

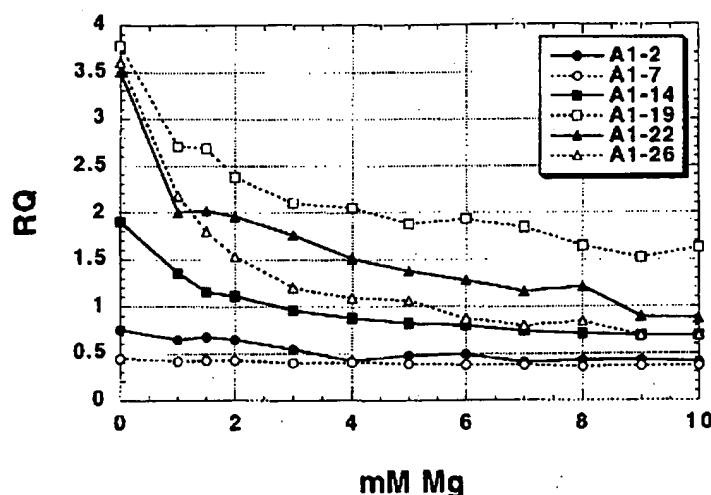
Regardless of the physical mechanism, the relative independence of position and quenching greatly simplifies the design of probes for the 5' nuclease PCR assay. There are three main factors that determine the performance of a double-labeled fluorescent probe in the 5' nuclease PCR assay. The first factor is the degree of quenching observed in the intact probe. This is characterized by the value of  $RQ^-$ , which is the ratio of reporter to quencher fluorescent emissions for a no template control PCR. Influences on the value of  $RQ^-$  include the particular reporter and quencher

**TABLE 3** Comparison of Fluorescence Emissions of Single-stranded and Double-stranded Fluorogenic Probes

Probe	518 nm		582 nm		RQ	
	ss	ds	ss	ds	ss	ds
A1-7	27.75	68.53	61.08	138.18	0.45	0.50
A1-26	43.31	509.38	53.50	93.86	0.81	5.43
A3-6	16.75	62.88	39.33	165.57	0.43	0.38
A3-24	30.05	578.64	67.72	140.25	0.45	3.21
P2-7	35.02	70.13	54.63	121.09	0.64	0.58
P2-27	39.89	320.47	65.10	61.13	0.61	5.25
P5-10	27.34	144.85	61.95	165.54	0.44	0.87
P5-28	33.65	462.29	72.39	104.61	0.46	4.43

(ss) Single-stranded. The fluorescence emissions at 518 or 582 nm for solutions containing a final concentration of 50 nM indicated probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 10 mM  $MgCl_2$ . (ds) Double-stranded. The solutions contained, in addition, 100 nM A1C for probes A1-7 and A1-26, 100 nM A3C for probes A3-6 and A3-24, 100 nM P2C for probes P2-7 and P2-27, or 100 nM P5C for probes P5-10 and P5-28. Before the addition of  $MgCl_2$ , 120  $\mu$ l of each sample was heated at 95°C for 5 min. Following the addition of 80  $\mu$ l of 25 mM  $MgCl_2$ , each sample was allowed to cool to room temperature and the fluorescence emissions were measured. Reported values are the average of three determinations.





**FIGURE 3** Effect of  $Mg^{2+}$  concentration on RQ ratio for the A1 series of probes. The fluorescence emission intensity at 518 and 582 nm was measured for solutions containing 50 nM probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and varying amounts (0–10 mM) of  $MgCl_2$ . The calculated RQ ratios (518 nm intensity divided by 582 nm intensity) are plotted vs.  $MgCl_2$  concentration (mM Mg). The key (upper right) shows the probes examined.

dyes used, spacing between reporter and quencher dyes, nucleotide sequence context effects, presence of structure or other factors that reduce flexibility of the oligonucleotide, and purity of the probe. The second factor is the efficiency of hybridization, which depends on probe  $T_m$ , presence of secondary structure in probe or template, annealing temperature, and other reaction conditions. The third factor is the efficiency at which *Taq* DNA polymerase cleaves the bound probe between the reporter and quencher dyes. This cleavage is dependent on sequence complementarity between probe and template as shown by the observation that mismatches in the segment between reporter and quencher dyes drastically reduce the cleavage of probe.<sup>(1)</sup>

The rise in  $RQ^-$  values for the A1 series of probes seems to indicate that the degree of quenching is reduced somewhat as the quencher is placed toward the 3' end. The lowest apparent quenching is observed for probe A1-19 (see Fig. 3) rather than for the probe where the TAMRA is at the 3' end (A1-26). This is understandable, as the conformation of the 3' end position would be expected to be less restricted than the conformation of an internal position. In effect, a quencher at the 3' end is freer to adopt conformations close to the 5' reporter dye than is an internally placed quencher. For the other three sets of

probes, the interpretation of  $RQ^-$  values is less clear-cut. The A3 probes show the same trend as A1, with the 3' TAMRA probe having a larger  $RQ^-$  than the internal TAMRA probe. For the P2 pair, both probes have about the same  $RQ^-$  value. For the P5 probes, the  $RQ^-$  for the 3' probe is less than for the internally labeled probe. Another factor that may explain some of the observed variation is that purity affects the  $RQ^-$  value. Although all probes are HPLC purified, a small amount of contamination with unquenched reporter can have a large effect on  $RQ^-$ .

Although there may be a modest effect on degree of quenching, the position of the quencher apparently can have a large effect on the efficiency of probe cleavage. The most drastic effect is observed with probe A1-2, where placement of the TAMRA on the second nucleotide reduces the efficiency of cleavage to almost zero. For the A3, P2, and P5 probes,  $\Delta RQ$  is much greater for the 3' TAMRA probes as compared with the internal TAMRA probes. This is explained most easily by assuming that probes with TAMRA at the 3' end are more likely to be cleaved between reporter and quencher than are probes with TAMRA attached internally. For the A1 probes, the cleavage efficiency of probe A1-7 must already be quite high, as  $\Delta RQ$  does not increase when the quencher is placed closer to the 3' end. This illus-

trates the importance of being able to use probes with a quencher on the 3' end in the 5' nuclease PCR assay. In this assay, an increase in the intensity of reporter fluorescence is observed only when the probe is cleaved between the reporter and quencher dyes. By placing the reporter and quencher dyes on the opposite ends of an oligonucleotide probe, any cleavage that occurs will be detected. When the quencher is attached to an internal nucleotide, sometimes the probe works well (A1-7) and other times not so well (A3-6). The relatively poor performance of probe A3-6 presumably means the probe is being cleaved 3' to the quencher rather than between the reporter and quencher. Therefore, the best chance of having a probe that reliably detects accumulation of PCR product in the 5' nuclease PCR assay is to use a probe with the reporter and quencher dyes on opposite ends.

Placing the quencher dye on the 3' end may also provide a slight benefit in terms of hybridization efficiency. The presence of a quencher attached to an internal nucleotide might be expected to disrupt base-pairing and reduce the  $T_m$  of a probe. In fact, a 2°C–3°C reduction in  $T_m$  has been observed for two probes with internally attached TAMRAs.<sup>(9)</sup> This disruptive effect would be minimized by placing the quencher at the 3' end. Thus, probes with 3' quenchers might exhibit slightly higher hybridization efficiencies than probes with internal quenchers.

The combination of increased cleavage and hybridization efficiencies means that probes with 3' quenchers probably will be more tolerant of mismatches between probe and target as compared with internally labeled probes. This tolerance of mismatches can be advantageous, as when trying to use a single probe to detect PCR-amplified products from samples of different species. Also, it means that cleavage of probe during PCR is less sensitive to alterations in annealing temperature or other reaction conditions. The one application where tolerance of mismatches may be a disadvantage is for allelic discrimination. Lee et al.<sup>(11)</sup> demonstrated that allele-specific probes were cleaved between reporter and quencher only when hybridized to a perfectly complementary target. This allowed them to distinguish the normal human cystic fibrosis allele from the  $\Delta F508$  mutant. Their probes had TAMRA attached to the seventh nucleotide from

# Research

the 5' end and were designed so that any mismatches were between the reporter and quencher. Increasing the distance between reporter and quencher would lessen the disruptive effect of mismatches and allow cleavage of the probe on the incorrect target. Thus, probes with a quencher attached to an internal nucleotide may still be useful for allelic discrimination.

In this study loss of quenching upon hybridization was used to show that quenching by a 3' TAMRA is dependent on the flexibility of a single-stranded oligonucleotide. The increase in reporter fluorescence intensity, though, could also be used to determine whether hybridization has occurred or not. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends should also be useful as hybridization probes. The ability to detect hybridization in real time means that these probes could be used to measure hybridization kinetics. Also, this type of probe could be used to develop homogeneous hybridization assays for diagnostics or other applications. Bagwell et al.<sup>(10)</sup> describe just this type of homogeneous assay where hybridization of a probe causes an increase in fluorescence caused by a loss of quenching. However, they utilized a complex probe design that requires adding nucleotides to both ends of the probe sequence to form two imperfect hairpins. The results presented here demonstrate that the simple addition of a reporter dye to one end of an oligonucleotide and a quencher dye to the other end generates a fluorogenic probe that can detect hybridization or PCR amplification.

## ACKNOWLEDGMENTS

We acknowledge Lincoln McBride of Perkin-Elmer for his support and encouragement on this project and Mitch Winnik of the University of Toronto for helpful discussions on time-resolved fluorescence.

## REFERENCES

1. Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* **21**: 3761-3766.
2. Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction prod-

uct by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* **88**: 7276-7280.

3. Lyamichev, V., M.A.D. Brow, and J.E. Dahlberg. 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* **260**: 778-783.
4. Förster, V.Th. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys. (Leipzig)* **2**: 55-75.
5. Lakowicz, J.R. 1983. Energy transfer. In *Principles of fluorescent spectroscopy*, pp. 303-339. Plenum Press, New York, NY.
6. Stryer, L. and R.P. Haugland. 1967. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci.* **58**: 719-726.
7. Nakajima-Iijima, S., H. Hamada, P. Reddy, and T. Kakunaga. 1985. Molecular structure of the human cytoplasmic beta-actin gene: Inter-species homology of sequences in the introns. *Proc. Natl. Acad. Sci.* **82**: 6133-6137.
8. du Breuil, R.M., J.M. Patel, and B.V. Mendelow. 1993. Quantitation of  $\beta$ -actin-specific mRNA transcripts using xeno-competitive PCR. *PCR Methods Applic.* **3**: 57-59.
9. Livak, K.J. (unpubl.).
10. Bagwell, C.B., M.E. Munson, R.L. Christensen, and E.J. Lovett. 1994. A new homogeneous assay system for specific nucleic acid sequences: Poly-dA and poly-A detection. *Nucleic Acids Res.* **22**: 2424-2425.

Received December 20, 1994; accepted in revised form March 6, 1995.



THIS MATERIAL MAY BE PROTECTED  
BY COPYRIGHT LAW (17 U.S. CODE)

## GENOME METHODS

## Real Time Quantitative PCR

Christian A. Heid,<sup>1</sup> Junko Stevens,<sup>2</sup> Kenneth J. Livak,<sup>2</sup> and  
P. Mickey Williams<sup>1,3</sup>

<sup>1</sup>BioAnalytical Technology Department, Genentech, Inc., South San Francisco, California 94080;

<sup>2</sup>Applied BioSystems Division of Perkin Elmer Corp., Foster City, California 94404

We have developed a novel "real time" quantitative PCR method. The method measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan Probe). This method provides very accurate and reproducible quantitation of gene copies. Unlike other quantitative PCR methods, real-time PCR does not require post-PCR sample handling, preventing potential PCR product carry-over contamination and resulting in much faster and higher throughput assays. The real-time PCR method has a very large dynamic range of starting target molecule determination (at least five orders of magnitude). Real-time quantitative PCR is extremely accurate and less labor-intensive than current quantitative PCR methods.

Quantitative nucleic acid sequence analysis has had an important role in many fields of biological research. Measurement of gene expression (RNA) has been used extensively in monitoring biological responses to various stimuli (Tan et al. 1994; Huang et al. 1995a,b; Prud'homme et al. 1995). Quantitative gene analysis (DNA) has been used to determine the genome quantity of a particular gene, as in the case of the human *HER2* gene, which is amplified in ~30% of breast tumors (Slamon et al. 1987). Gene and genome quantitation (DNA and RNA) also have been used for analysis of human immunodeficiency virus (HIV) burden demonstrating changes in the levels of virus throughout the different phases of the disease (Connor et al. 1993; Platak et al. 1993b; Furtado et al. 1995).

Many methods have been described for the quantitative analysis of nucleic acid sequences (both for RNA and DNA; Southern 1975; Sharp et al. 1980; Thomas 1980). Recently, PCR has proven to be a powerful tool for quantitative nucleic acid analysis. PCR and reverse transcriptase (RT)-PCR have permitted the analysis of minimal starting quantities of nucleic acid (as little as one cell equivalent). This has made possible many experiments that could not have been performed with traditional methods. Although PCR has provided a powerful tool, it is imperative

that it be used properly for quantitation (Rasmackers 1995). Many early reports of quantitative PCR and RT-PCR described quantitation of the PCR product but did not measure the initial target sequence quantity. It is essential to design proper controls for the quantitation of the initial target sequences (Perre 1992; Clementi et al. 1993).

Researchers have developed several methods of quantitative PCR and RT-PCR. One approach measures PCR product quantity in the log phase of the reaction before the plateau (Kellogg et al. 1990; Pang et al. 1990). This method requires that each sample has equal input amounts of nucleic acid and that each sample under analysis amplifies with identical efficiency up to the point of quantitative analysis. A gene sequence (contained in all samples at relatively constant quantities, such as  $\beta$ -actin) can be used for sample amplification efficiency normalization. Using conventional methods of PCR detection and quantitation (gel electrophoresis or plate capture hybridization), it is extremely laborious to assure that all samples are analyzed during the log phase of the reaction (for both the target gene and the normalization gene). Another method, quantitative competitive (QC)-PCR, has been developed and is used widely for PCR quantitation. QC-PCR relies on the inclusion of an internal control competitor in each reaction (Becker-Andre 1991; Ptacek et al. 1993a,b). The efficiency of each reaction is normalized to the internal competitor. A known amount of internal competitor can be

<sup>3</sup>Corresponding author.

## REAL TIME QUANTITATIVE PCR

## RESULTS

## PCR Product Detection in Real Time

added to each sample. To obtain relative quantitation, the unknown target PCR product is compared with the known competitor PCR product. Success of a quantitative competitive PCR assay relies on developing an internal control that amplifies with the same efficiency as the target molecule. The design of the competitor and the validation of amplification efficiencies require a dedicated effort. However, because QC-PCR does not require that PCR products be analyzed during the log phase of the amplification, it is the easier of the two methods to use.

Several detection systems are used for quantitative PCR and RT-PCR analysis: (1) agarose gels, (2) fluorescent labelling of PCR products and detection with laser-induced fluorescence using capillary electrophoresis (Fusco et al. 1995; Williams et al. 1996) or acrylamide gels, and (3) plate capture and sandwich probe hybridization (Mulder et al. 1994). Although these methods proved successful, each method requires post-PCR manipulations that add time to the analysis and may lead to laboratory contamination. The sample throughput of these methods is limited (with the exception of the plate capture approach), and, therefore, these methods are not well suited for uses demanding high sample throughput (i.e., screening of large numbers of biomolecules or analyzing samples for diagnostics or clinical trials).

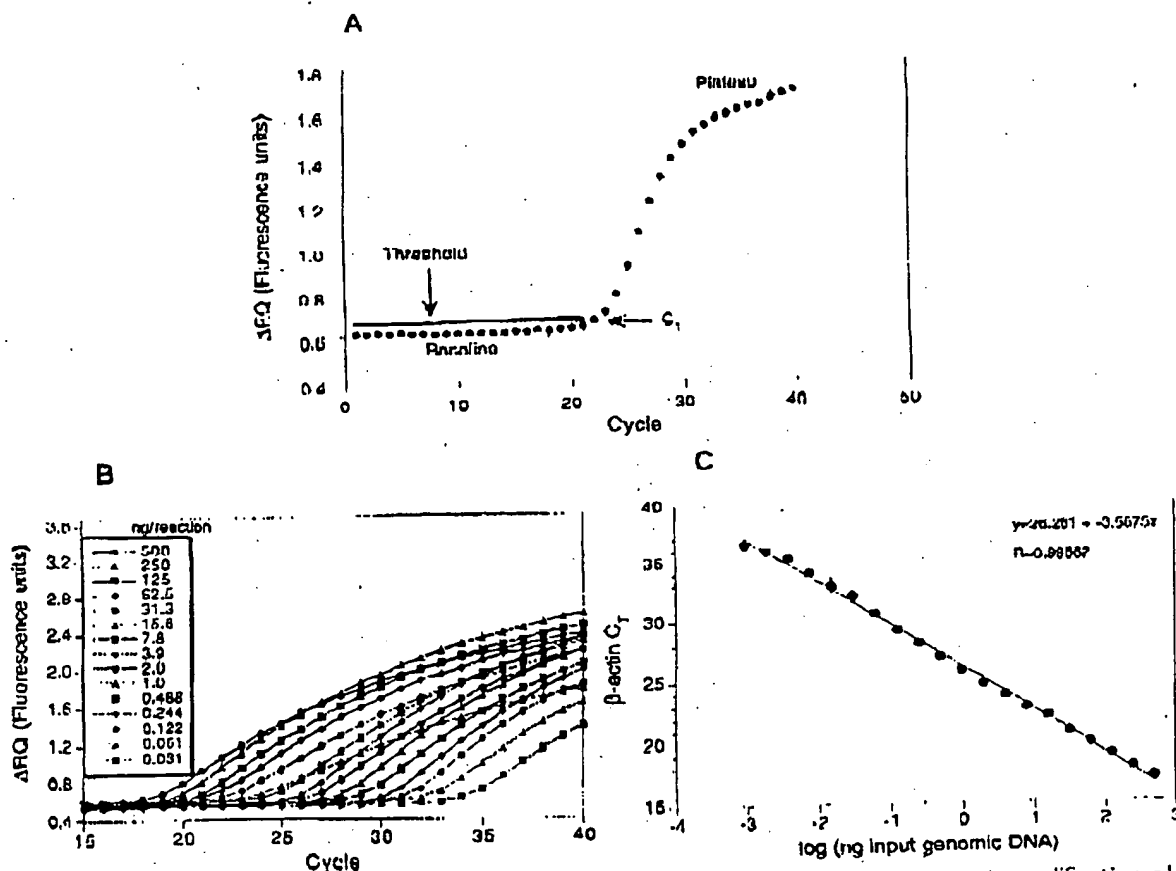
Here we report the development of a novel assay for quantitative DNA analysis. The assay is based on the use of the 5' nuclease assay first described by Holland et al. (1991). The method uses the 5' nuclease activity of *Taq* polymerase to cleave a nonextendible hybridization probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee et al. 1993; Bassler et al. 1995; Livak et al. 1995a,b). One fluorescent dye serves as a reporter [FAM (i.e., 6-carboxyfluorescein)] and its emission spectra is quenched by the second fluorescent dye, TAMRA (i.e., 6-carboxy-tetramethylrhodamine). The nuclease degradation of the hybridization probe releases the quenching of the FAM fluorescent emission, resulting in an increase in peak fluorescent emission at 518 nm. The use of a sequence detector (ABI Prism) allows measurement of fluorescent spectra of all 96 wells of the thermal cycler continuously during the PCR amplification. Therefore, the reactions are monitored in real time. The output data is described and quantitative analysis of input target DNA sequences is discussed below.

The goal was to develop a high-throughput, sensitive, and accurate gene quantitation assay for use in monitoring lipid mediated therapeutic gene delivery. A plasmid encoding human factor VIII gene sequence, pF8TM (see Methods), was used as a model therapeutic gene. The assay uses fluorescent Taqman methodology and an instrument capable of measuring fluorescence in real time (ABI Prism 7700 Sequence Detector). The Taqman reaction requires a hybridization probe labeled with two different fluorescent dyes. One dye is a reporter dye (FAM), the other is a quenching dye (TAMRA). When the probe is intact, fluorescent energy transfer occurs and the reporter dye fluorescent emission is absorbed by the quenching dye (TAMRA). During the extension phase of the PCR cycle, the fluorescent hybridization probe is cleaved by the 5'-3' nucleolytic activity of the DNA polymerase. On cleavage of the probe, the reporter dye emission is no longer transferred efficiently to the quenching dye, resulting in an increase of the reporter dye fluorescent emission spectra. PCR primers and probes were designed for the human factor VIII sequence and human  $\beta$ -actin gene (as described in Methods). Optimization reactions were performed to choose the appropriate probe and magnesium concentrations yielding the highest intensity of reporter fluorescent signal without sacrificing specificity. The instrument uses a charge-coupled device (i.e., CCD camera) for measuring the fluorescent emission spectra from 500 to 650 nm. Each PCR tube was monitored sequentially for 25 msec with continuous monitoring throughout the amplification. Each tube was re-examined every 8.5 sec. Computer software was designed to examine the fluorescent intensity of both the reporter dye (FAM) and the quenching dye (TAMRA). The fluorescent intensity of the quenching dye, TAMRA, changes very little over the course of the PCR amplification (data not shown). Therefore, the intensity of TAMRA dye emission serves as an internal standard with which to normalize the reporter dye (FAM) emission variations. The software calculates a value termed  $\Delta Rn$  (or  $\Delta R(2)$ ) using the following equation:  $\Delta Rn = (Rn^i) - (Rn^f)$ , where  $Rn^i$  = emission intensity of reporter/emission intensity of quencher at any given time in a reaction tube, and  $Rn^f$  = emission intensity of re-

## HUID ET AL.

porter/emission intensity of quencher measured prior to PCR amplification in that same reaction tube. For the purpose of quantitation, the last three data points ( $\Delta Rn$ s) collected during the extension step for each PCR cycle were analyzed. The nucleolytic degradation of the hybridization probe occurs during the extension phase of PCR, and, therefore, reporter fluorescent emission increases during this time. The three data points were averaged for each PCR cycle and the mean value for each was plotted in an "amplification plot" shown in Figure 1A. The  $\Delta Rn$  mean value is plotted on the y-axis, and time, represented by cycle number, is plotted on the x-axis. During the early cycles of the PCR amplification, the  $\Delta Rn$

value remains at base line. When sufficient hybridization probe has been cleaved by the *Taq* polymerase nuclease activity, the intensity of reporter fluorescent emission increases. Most PCR amplifications reach a plateau phase of reporter fluorescent emission if the reaction is carried out to high cycle numbers. The amplification plot is examined early in the reaction, at a point that represents the log phase of product accumulation. This is done by assigning an arbitrary threshold that is based on the variability of the base-line data. In Figure 1A, the threshold was set at 10 standard deviations above the mean of base line emission calculated from cycles 1 to 15. Once the threshold is chosen, the point at which



**Figure 1** PCR product detection in real time. (A) The Model 7700 software will construct amplification plots from the extension phase fluorescent emission data collected during the PCR amplification. The standard deviation is determined from the data points collected from the base line of the amplification plot.  $C_T$  values are calculated by determining the point at which the fluorescence exceeds a threshold limit (usually 10 times the standard deviation of the base line). (B) Overlay of amplification plots of serially (1:2) diluted human genomic DNA samples amplified with  $\beta$ -actin primers. (C) Input DNA concentration of the samples plotted versus  $C_T$ . All

## REAL TIME QUANTITATIVE PCR

the amplification plot crosses the threshold is defined as  $C_T$ .  $C_T$  is reported as the cycle number at this point. As will be demonstrated, the  $C_T$  value is predictive of the quantity of input target.

### $C_T$ Values Provide a Quantitative Measurement of Input Target Sequences

Figure 1B shows amplification plots of 15 different PCR amplifications overlaid. The amplifications were performed on a 1:2 serial dilution of human genomic DNA. The amplified target was human  $\beta$  actin. The amplification plots shift to the right (to higher threshold cycles) as the input target quantity is reduced. This is expected because reactions with fewer starting copies of the target molecule require greater amplification to degrade enough probe to attain the threshold fluorescence. An arbitrary threshold of 10 standard deviations above the base line was used to determine the  $C_T$  values. Figure 1C represents the  $C_T$  values plotted versus the sample dilution value. Each dilution was amplified in triplicate PCR amplifications and plotted as mean values with error bars representing one standard deviation. The  $C_T$  values decrease linearly with increasing target quantity. Thus,  $C_T$  values can be used as a quantitative measurement of the input target number. It should be noted that the amplification plot for the 15.6-ng sample shown in Figure 1B does not reflect the same fluorescent rate of increase exhibited by most of the other samples. The 15.6-ng sample also achieves endpoint plateau at a lower fluorescent value than would be expected based on the input DNA. This phenomenon has been observed occasionally with other samples (data not shown) and may be attributable to late cycle inhibition; this hypothesis is still under investigation. It is important to note that the flattened slope and early plateau do not impact significantly the calculated  $C_T$  value as demonstrated by the fit on the line shown in Figure 1C. All triplicate amplifications resulted in very similar  $C_T$  values—the standard deviation did not exceed 0.5 for any dilution. This experiment contains a >100,000-fold range of input target molecules. Using  $C_T$  values for quantitation permits a much larger assay range than directly using total fluorescent emission intensity for quantitation. The linear range of fluorescent intensity measurement of the ABI Prism 7700 Se-

ments over a very large range of relative starting target quantities.

### Sample Preparation Validation

Several parameters influence the efficiency of PCR amplification: magnesium and salt concentrations, reaction conditions (i.e., time and temperature), PCR target size and composition, primer sequences, and sample purity. All of the above factors are common to a single PCR assay, except sample to sample purity. In an effort to validate the method of sample preparation for the factor VIII assay, PCR amplification reproducibility and efficiency of 10 replicate sample preparations were examined. After genomic DNA was prepared from the 10 replicate samples, the DNA was quantitated by ultraviolet spectroscopy. Amplifications were performed analyzing  $\beta$ -actin gene content in 100 and 25 ng of total genomic DNA. Each PCR amplification was performed in triplicate. Comparison of  $C_T$  values for each triplicate sample show minimal variation based on standard deviation and coefficient of variance (Table 1). Therefore, each of the triplicate PCR amplifications was highly reproducible, demonstrating that real time PCR using this instrumentation introduces minimal variation into the quantitative PCR analysis. Comparison of the mean  $C_T$  values of the 10 replicate sample preparations also showed minimal variability, indicating that each sample preparation yielded similar results for  $\beta$ -actin gene quantity. The highest  $C_T$  difference between any of the samples was 0.85 and 0.71 for the 100 and 25 ng samples, respectively. Additionally, the amplification of each sample exhibited an equivalent rate of fluorescent emission intensity change per amount of DNA target analyzed as indicated by similar slopes derived from the sample dilutions (Fig. 2). Any sample containing an excess of a PCR inhibitor would exhibit a greater measured  $\beta$ -actin  $C_T$  value for a given quantity of DNA. In addition, the inhibitor would be diluted along with the sample in the dilution analysis (Fig. 2), altering the expected  $C_T$  value change. Each sample amplification yielded a similar result in the analysis, demonstrating that this method of sample preparation is highly reproducible with regard to sample purity.

### Quantitative Analysis of a Plasmid After

PCR 001 RFR VVJ RC:BT 7007/CO/7T

## FINAL

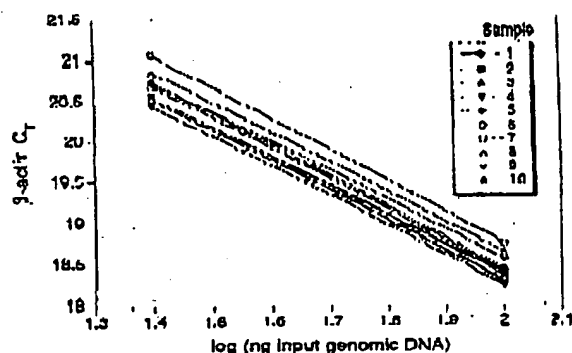
Table 1. Reproducibility of Sample Preparation Method

Sample no.	100 ng				25 ng			
	C <sub>T</sub>	mean	standard deviation	CV	C <sub>T</sub>	mean	standard deviation	CV
1	18.24	18.27	0.06	0.32	20.48	20.51	0.03	0.17
	18.23				20.55			
	18.33				20.5			
2	18.33	18.37	0.06	0.32	20.61	20.54	0.11	0.54
	18.35				20.59			
	18.44				20.41			
3	18.3	18.34	0.07	0.36	20.54	20.54	0.06	0.28
	18.3				20.6			
	18.42				20.49			
4	18.15	18.23	0.08	0.46	20.48	20.43	0.05	0.26
	18.23				20.44			
	18.32				20.38			
5	18.4	18.42	0.04	0.23	20.68	20.73	0.13	0.61
	18.38				20.87			
	18.46				20.63			
6	18.54	18.74	0.24	1.26	21.09	21.06	0.03	0.15
	18.67				21.04			
	19				21.04			
7	18.28	18.39	0.12	0.66	20.67	20.68	0.04	0.2
	18.36				20.73			
	18.52				20.65			
8	18.45	18.63	0.16	0.83	20.98	20.86	0.12	0.57
	18.7				20.84			
	18.73				20.75			
9	18.18	18.29	0.1	0.55	20.46	20.51	0.07	0.32
	18.34				20.54			
	18.26				20.48			
10	18.42	18.55	0.12	0.65	20.79	20.73	0.1	0.16
	18.57				20.78			
	18.66				20.62			
Mean	(1 10)	18.42	0.17	0.90		20.66	0.19	0.94

for containing a partial cDNA for human factor VIII, pF8TM. A series of transfections was set up using a decreasing amount of the plasmid (40, 4, 0.5, and 0.1 µg). Twenty-four hours post-transfection, total DNA was purified from each flask of cells. β-Actin gene quantity was chosen as a value for normalization of genomic DNA concentration from each sample. In this experiment, β-actin gene content should remain constant relative to total genomic DNA. Figure 3 shows the result of the β-actin DNA measurement (100 ng total DNA determined by ultraviolet spectroscopy) of each sample. Each sample was analyzed in triplicate and the mean β-actin C<sub>T</sub> values of the triplicates were plotted (error bars represent standard deviation). The highest difference

between any two sample means was 0.95 C<sub>T</sub>. Ten nanograms of total DNA of each sample were also examined for β-actin. The results again showed that very similar amounts of genomic DNA were present; the maximum mean β-actin C<sub>T</sub> value difference was 1.0. As figure 3 shows, the rate of β-actin C<sub>T</sub> change between the 100 and 10-ng samples was similar (slope values range between -3.56 and -3.45). This verifies again that the method of sample preparation yields samples of identical PCR integrity (i.e., no sample contained an excessive amount of a PCR inhibitor). However, these results indicate that each sample contained slight differences in the actual amount of genomic DNA analyzed. Determination of actual genomic DNA concentration was accomplished

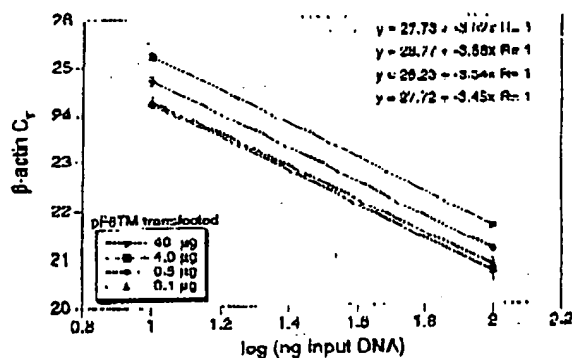
## REAL TIME- QUANTITATIVE PCR



**Figure 2** Sample preparation purity. The replicate samples shown in Table 1 were also amplified in triplicate using 25 ng of each DNA sample. The figure shows the input DNA concentration (100 and 25 ng) vs.  $C_T$ . In the figure, the 100 and 25 ng points for each sample are connected by a line.

by plotting the mean  $\beta$ -actin  $C_T$  value obtained for each 100-ng sample on a  $\beta$ -actin standard curve (shown in Fig. 4C). The actual genomic DNA concentration of each sample,  $a$ , was obtained by extrapolation to the x-axis.

Figure 4A shows the measured (i.e., non-normalized) quantities of factor VIII plasmid DNA (pF8TM) from each of the four transient cell transfections. Each reaction contained 100 ng of total sample DNA (as determined by UV spectroscopy). Each sample was analyzed in triplicate



**Figure 3** Analysis of transfected cell DNA quantity and purity. The DNA preparations of the four 293 cell transfections (40, 4, 0.5, and 0.1  $\mu$ g of pF8TM) were analyzed for the  $\beta$ -actin gene. 100 and 10 ng (determined by ultraviolet spectroscopy) of each sample were amplified in triplicate. For each amount of pF8TM that was transfected, the  $\beta$ -actin  $C_T$  values are plotted versus the total input DNA concentration.

PCR amplifications. As shown, pF8TM purified from the 293 cells decreases (mean  $C_T$  values increase) with decreasing amounts of plasmid transfected. The mean  $C_T$  values obtained for pF8TM in Figure 4A were plotted on a standard curve comprised of serially diluted pF8TM, shown in Figure 4B. The quantity of pF8TM,  $b$ , found in each of the four transfections was determined by extrapolation to the x-axis of the standard curve in Figure 4B. These uncorrected values,  $b$ , for pF8TM were normalized to determine the actual amount of pF8TM found per 100 ng of genomic DNA by using the equation:

$$\frac{b \times 100 \text{ ng}}{a} = \text{actual pF8TM copies per 100 ng of genomic DNA}$$

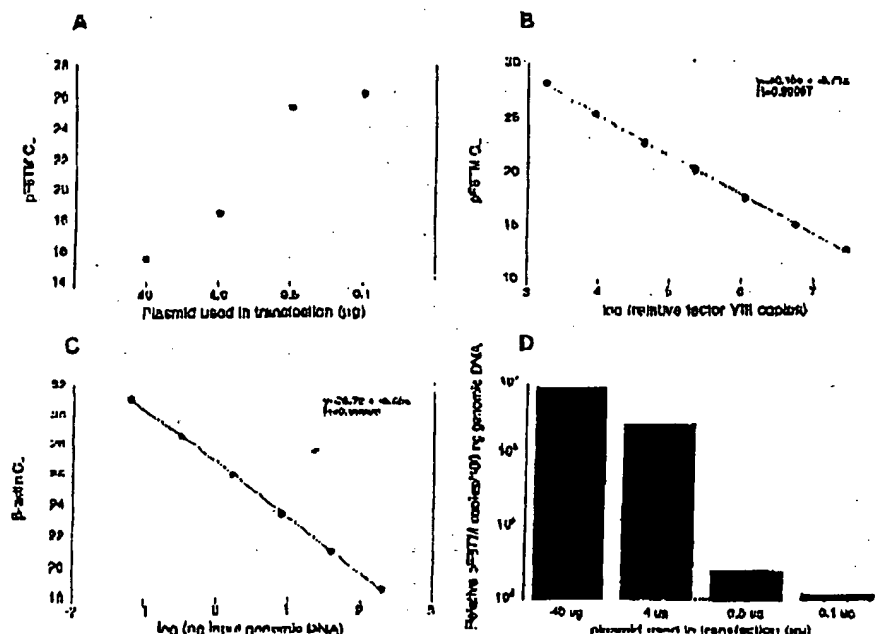
where  $a$  = actual genomic DNA in a sample and  $b$  = pF8TM copies from the standard curve. The normalized quantity of pF8TM per 100 ng of genomic DNA for each of the four transfections is shown in Figure 4D. These results show that the quantity of factor VIII plasmid associated with the 293 cells, 24 hr after transfection, decreases with decreasing plasmid concentration used in the transfection. The quantity of pF8TM associated with 293 cells, after transfection with 40  $\mu$ g of plasmid, was 35  $\mu$ g per 100 ng genomic DNA. This results in ~320 plasmid copies per cell.

## DISCUSSION

We have described a new method for quantitating gene copy numbers using real-time analysis of PCR amplifications. Real-time PCR is compatible with either of the two PCR (RT-PCR) approaches: (1) quantitative competitive where an internal competitor for each target sequence is used for normalization (data not shown) or (2) quantitative comparative PCR using a normalization gene contained within the sample (i.e.,  $\beta$ -actin) or a "housekeeping" gene for RT-PCR. If equal amounts of nucleic acid are analyzed for each sample and if the amplification efficiency before quantitative analysis is identical for each sample, the internal control (normalization gene or competitor) should give equal signals for all samples.

The real-time PCR method offers several advantages over the other two methods currently employed (see the Introduction). First, the real-time PCR method is performed in a closed-tube system and requires no post-PCR manipulation

HUID LI AL.



**Figure 4** Quantitative analysis of pF8TM in transfected cells. (A) Amount of plasmid DNA used for the transfection plotted against the mean  $C_i$  value determined for pF8TM remaining 24 hr after transfection. (B,C) Standard curves of pF8TM and  $\beta$ -actin, respectively. pF8TM DNA (B) and genomic DNA (C) were diluted serially 1:5 before amplification with the appropriate primers. The  $\beta$ -actin standard curve was used to normalize the results of A to 100 ng of genomic DNA. (D) The amount of pF8TM present per 100 ng of genomic DNA.

of sample. Therefore, the potential for PCR contamination in the laboratory is reduced because amplified products can be analyzed and disposed of without opening the reaction tubes. Second, this method supports the use of a normalization gene (i.e.,  $\beta$ -actin) for quantitative PCR or housekeeping genes for quantitative RT-PCR controls. Analysis is performed in real time during the log phase of product accumulation. Analysis during log phase permits many different genes (over a wide input target range) to be analyzed simultaneously, without concern of reaching reaction plateau at different cycles. This will make multi-gene analysis assays much easier to develop, because individual internal competitors will not be needed for each gene under analysis. Third, sample throughput will increase dramatically with the new method because there is no post-PCR processing time. Additionally, working in a 96-well format is highly compatible with automation technology.

The real-time PCR method is highly reproducible. Replicate amplifications can be analyzed

for each sample minimizing potential error. The system allows for a very large assay dynamic range (approaching 1,000,000-fold starting target). Using a standard curve for the target of interest, relative copy number values can be determined for any unknown sample. Fluorescent threshold values,  $C_{\text{p}}$ , correlate linearly with relative DNA copy numbers. Real time quantitative RT-PCR methodology (Gibson et al., this issue) has also been developed. Finally, real time quantitative PCR methodology can be used to develop high-throughput screening assays for a variety of applications [quantitative gene expression (RT-PCR), gene copy assays (Her2, IIIV, etc.), genotyping (knockout mouse analysis), and immuno-PCR].

Real-time PCR may also be performed using intercalating dyes (Higuchi et al. 1992) such as ethidium bromide. The fluorogenic probe method offers a major advantage over intercalating dyes—greater specificity (i.e., primer dimers and nonspecific PCR products are not detected).

## METHODS

### Generation of a Plasmid Containing a Partial cDNA for Human Factor VIII

Total RNA was harvested (RNAzol B from Tel Test, Inc., Friendswood, TX) from cells transfected with a factor VIII expression vector, pCIS2.8c251 (Eaton et al. 1986; Gorman et al. 1990). A factor VIII partial cDNA sequence was generated by RT-PCR [CloneAmp II<sup>2</sup> with RNA PCR Kit (part N808-0179, PE Applied Biosystems, Foster City, CA)] using the PCR primers F8for and F8rev (primer sequences are shown below). The amplicon was reamplified using modified F8for and F8rev primers (appended with *Hind*III and *Hind*III restriction site sequences at the 5' end) and cloned into pCIS2.3Z (Promega Corp., Madison, WI). The resulting clone, pF8TM, was used for transient transfection of 293 cells.

### Amplification of Target DNA and Detection of Amplicon Factor VIII Plasmid DNA

(pF8TM) was amplified with the primers F8for 5'-CCGCTTCCCAAGAGTGAAGTTC-3' and F8rev 5'-AAACCTTCAACCTTGGATGCTAGG-3'. The reaction produced a 422-bp PCR product. The forward primer was designed to recognize a unique sequence found in the 5' untranslated region of the parent pCIS2.8c251 plasmid and therefore does not recognize and amplify the human factor VIII gene. Primers were chosen with the assistance of the computer program Oligo 4.0 (National Biosciences, Inc., Plymouth, MN). The human  $\beta$ -actin gene was amplified with the primers  $\beta$ -actin forward primer 5'-TCACCCACACTCTT GCCATCTTACCA-3' and  $\beta$ -actin reverse primer 5'-CAGC CGGAACCCCTTCATTGCCAATGG-3'. The reaction produced a 295-bp PCR product.

Amplification reactions (50  $\mu$ l) contained a DNA sample, 10X PCR Buffer II (5  $\mu$ l), 200  $\mu$ M dATP, dCTP, dGTP, and 400  $\mu$ M dTTP, 4 mM MgCl<sub>2</sub>, 1.25 Units AmpliTag DNA polymerase, 0.5 unit AmpErase uracil N-glycosylase (UNG), 60 pmole of each factor VIII primer, and 15 pmole of each  $\beta$ -actin primer. The reactions also contained one of the following detection probes (100 nM each): F8probe 5'-(FAM)AGCTTCTCCACCTTCTCTCTTCTCTTGCCTT(TAMRA)p 3' and  $\beta$ -actin probe 5'-(FAM)ATGCCCTX(TAMRA)CCCCCATGCCATCp-3' where p indicates phosphorylation and X indicates a linker arm nucleotide. Reaction tubes were MicroAmp Optical Tubes (part number N801 0933, Perkin Elmer) that were frosted (at Perkin Elmer) to prevent light from reflecting. Tube caps were similar to MicroAmp Caps but specially designed to prevent light scattering. All of the PCR consumables were supplied by PE Applied Biosystems (Foster City, CA) except the factor VIII primers, which were synthesized at Genentech, Inc. (South San Francisco, CA). Probes were designed using the Oligo 4.0 software, following guidelines suggested in the Model 7700 Sequence Detector instrument manual. Briefly, probe T<sub>m</sub> should be at least 5°C higher than the annealing temperature used during thermal cycling; primers should not form stable duplexes with the probe.

The thermal cycling conditions included 2 min at 50°C and 10 min at 95°C. Thermal cycling proceeded with

## REAL TIME QUANTITATIVE PCR

reactions were performed in the Model 7700 Sequence Detector (PE Applied Biosystems), which contains a GeneAmp PCR System 9600. Reaction conditions were programmed on a Power Macintosh 7100 (Apple Computer, Santa Clara, CA) linked directly to the Model 7700 Sequence Detector. Analysis of data was also performed on the Macintosh computer. Collection and analysis software was developed at PE Applied Biosystems.

### Transfection of Cells with Factor VIII Construct

Four T175 flasks of 293 cells (ATCC CRL 1573), a human fetal kidney suspension cell line, were grown to 80% confluency and transfected pF8TM. Cells were grown in the following media: 50% HAM'S #12 without GHT, 50% low glucose Dulbecco's modified Eagle medium (DMEM) without glycine with sodium bicarbonate, 10% fetal bovine serum, 2 mM L-glutamine, and 1% penicillin-streptomycin. The media was changed 30 min before the transfection. pF8TM DNA amounts of 40, 4, 0.5, and 0.1  $\mu$ g were added to 1.5 ml of a solution containing 0.125 M CaCl<sub>2</sub> and 1X HBPS. The four mixtures were left at room temperature for 10 min and then added dropwise to the cells. The flasks were incubated at 37°C and 5% CO<sub>2</sub> for 24 hr, washed with PBS, and resuspended in PBS. The remaining cells were divided into aliquots and DNA was extracted immediately using the QIAamp Blood Kit (Qiagen, Chatsworth, CA). DNA was eluted into 200  $\mu$ l of 20 mM Tris-HCl at pH 8.0.

## ACKNOWLEDGMENTS

We thank Genentech's DNA Synthesis Group for primer synthesis and Genentech's Graphics Group for assistance with the figures.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Bassler, H.A., S.J. Flood, K.J. Livak, J. Marimano, R. Kimm, and C.A. Ball. 1995. Use of a fluorogenic probe in a PCR-based assay for the detection of *Listeria monocytogenes*. *App. Environ. Microbiol.* 61: 3724-3728.
- Becker-Andre, M. 1991. Quantitative evaluation of mRNA levels. *Meth. Mol. Cell. Biol.* 2: 189-201.
- Clementi, M., S. Menzo, P. Baggiarelli, A. Manzoli, A. Valenza, and P.E. Varaldo. 1997. Quantitative PCR and RT-PCR in virology. [Review]. *PCR Methods Applic.* 2: 191-196.
- Connor, R.J., H. Mohi, Y. Cao, and D.D. Ho. 1993. Increased viral burden and cytopathicity correlate temporally with CD4<sup>+</sup> T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 67: 1772-1777.
- Eaton, D.L., W.J. Wood, D. Eaton, P.E. Hoss, P.



## HFID LI AL

Venar, and C. Gorman. 1986. Construction and characterization of an active factor VIII variant lacking the central one third of the molecule. *Biochemistry* 25: 8343-8347.

Fasco, M.J., C.P. Treanor, S. Spivack, H.L. Wigge, and L.S. Kaminsky. 1995. Quantitative RNA-polymerase chain reaction-DNA analysis by capillary electrophoresis and laser-induced fluorescence. *Anal. Biochem.* 224: 140-147.

Forre, J. 1992. Quantitative or semi-quantitative PCR: Reality versus myth. *PCR Methods Applic.* 2: 1-9.

Furtado, M.R., L.A. Kingsley, and S.M. Wollinsky. 1995. Changes in the viral mRNA expression pattern correlate with a rapid rate of CD4+ T-cell number decline in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 69: 2092-2100.

Gibson, U.E.M., C.A. Heid, and P.M. Williams. 1996. A novel method for real time quantitative competitive RT-PCR. *Genome Res.* (this issue).

Gorman, C.M., D.R. Giles, and G. McCray. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Prot. Engin. Tech.* 2: 3-10.

Higuchi, R., G. Dollinger, P.S. Walsh, and R. Griffith. 1992. Simultaneous amplification and detection of specific DNA sequences. *Biochemistry* 10: 413-417.

Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci.* 88: 7276-7280.

Huang, S.K., H.Q. Xiao, T.J. Klein, G. Paciotti, D.G. Marsh, L.M. Lichtenstein, and M.C. Liu. 1995a. IL-13 expression at the sites of allergen challenge in patients with asthma. *J. Immun.* 155: 2688-2694.

Huang, S.K., M. Yi, E. Palmer, and D.G. Marsh. 1995b. A dominant T cell receptor beta-chain in response to a short ragweed allergen. *Am J. Immun.* 154: 6157-6162.

Kellogg, D.E., J.J. Shtinsky, and S. Kowk. 1990. Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction. *Anal. Biochem.* 189: 202-208.

Lee, J.-G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761-3766.

Livak, K.J., S.J. Flood, J. Marmaro, W. Giusti, and K. Deetz. 1995a. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Applic.* 4: 357-362.

Livak, K.J., J. Marmaro, and J.A. Todd. 1995b. Towards

fully automated genome-wide polymorphism screening [Letter]. *Nature Genet.* 9: 341-342.

Mulder, J., N. McKinney, C. Christopherson, J. Shtinsky, L. Greenfield, and S. Kwak. 1994. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: Application to acute retroviral infection. *J. Clin. Microbiol.* 32: 292-300.

Pang, S., Y. Koyanagi, S. Miller, C. Wiloy, H.V. Vinters, and L.S. Chen. 1990. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* 343: 85-89.

Platak, M.J., K.C. Luk, B. Williams, and J.D. Lifson. 1993a. Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Aid Techniques* 14: 70-81.

Platak, M.J., M.S. Saag, L.C. Yang, S.J. Clark, J.C. Kappes, K.C. Luk, B.H. Hann, G.M. Shaw, and J.D. Lifson. 1993b. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR [see Comments]. *Science* 259: 1749-1754.

Prud'homme, G.J., D.H. Kono, and A.N. Theofilopoulos. 1995. Quantitative polymerase chain reaction analysis reveals marked overexpression of interleukin-1 beta, interleukin-1 and interferon-gamma mRNA in the lymph nodes of lupus-prone mice. *Mol. Immunol.* 32: 495-503.

Racynackers, L. 1995. A commentary on the practical applications of competitive PCR. *Genome Res.* 5: 91-94.

Sharp, P.A., A.J. Berk, and S.M. Berget. 1980. Transcription maps of adenovirus. *Methods Enzymol.* 65: 750-768.

Slamon, D.J., G.M. Clark, S.G. Wong, W.J. Levin, A. Ulrich, and W.L. McGuire. 1987. Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235: 177-182.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.

Tan, X., X. Sun, C.F. Gonzalez, and W. Hsueh. 1994. PAF and TNF increase the precursor of NF-kappa B p50 mRNA in mouse intestine: Quantitative analysis by competitive PCR. *Biochim. Biophys. Acta* 1215: 157-162.

Thomas, P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci.* 77: 5201-5205.

Williams, S., C. Schwer, A. Krishnarao, C. Heid, B. Karger, and P.M. Williams. 1996. Quantitative competitive PCR: Analysis of amplified products of the HIV-1 gag gene by capillary electrophoresis with laser induced fluorescence detection. *Anal. Biochem.* (in press).

Received June 3, 1996; accepted in revised form July 29, 1996.

## WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA<sup>\*†</sup>, TODD A. SWANSON<sup>\*</sup>, JAMES W. WELSH<sup>\*</sup>, MARGARET A. ROY<sup>‡</sup>, DAVID A. LAWRENCE<sup>\*</sup>, JAMES LEE<sup>‡</sup>, JENNIFER BRUSH<sup>‡</sup>, LISA A. TANEYHILL<sup>§</sup>, BETHANNE DEUEL<sup>‡</sup>, MICHAEL LEW<sup>¶</sup>, COLIN WATANABE<sup>¶</sup>, ROBERT L. COHEN<sup>\*</sup>, MONA F. MELHEM<sup>\*\*</sup>, GENE G. FINLEY<sup>\*\*</sup>, PHIL QUIRKE<sup>††</sup>, AUDREY D. GODDARD<sup>‡</sup>, KENNETH J. HILLAN<sup>¶</sup>, AUSTIN L. GURNEY<sup>‡</sup>, DAVID BOTSTEIN<sup>‡,‡‡</sup>, AND ARNOLD J. LEVINE<sup>§</sup>

Departments of <sup>\*</sup>Molecular Oncology, <sup>‡</sup>Molecular Biology, <sup>§</sup>Scientific Computing, and <sup>¶</sup>Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; <sup>\*\*</sup>University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; <sup>††</sup>University of Leeds, Leeds, LS29JT United Kingdom; <sup>‡‡</sup>Department of Genetics, Stanford University, Palo Alto, CA 94305; and <sup>§</sup>Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

**ABSTRACT** Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ) resulting in an increase in  $\beta$ -catenin levels. Stabilized  $\beta$ -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating  $\beta$ -catenin levels (9). APC is phosphorylated by GSK-3 $\beta$ , binds to  $\beta$ -catenin, and facilitates its degradation. Mutations in either APC or  $\beta$ -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- $\beta$  superfamily, and the homeobox genes, *engrailed*, *goosecoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

### MATERIALS AND METHODS

**SSH.** SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

To whom reprint requests should be addressed. e-mail: diane@gene.com.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-6\$2.00/0  
PNAS is available online at www.pnas.org.

cDNA was synthesized from 2  $\mu$ g of poly(A)<sup>+</sup> RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2  $\mu$ g of poly(A)<sup>+</sup> RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

**cDNA Library Screening.** Clones encoding full-length mouse *WISP-1* were isolated by screening a  $\lambda$ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening  $\lambda$ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

**Expression of Human *WISP* RNA.** PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300  $\mu$ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

**In Situ Hybridization.** <sup>33</sup>P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

**Radiation Hybrid Mapping.** Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

**Cell Lines, Tumors, and Mucosa Specimens.** Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNazol.

**Gene Amplification and RNA Expression Analysis.** Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula  $2^{-\Delta Ct}$  where  $\Delta Ct$  represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The  $\delta$ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

## RESULTS

**Isolation of *WISP-1* and *WISP-2* by SSH.** To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1A and B). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on  $\beta$ -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of  $\approx 40,000$  ( $M_r$  40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2A).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of  $\approx 27,000$  ( $M_r$  27 K) (Fig. 2B). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

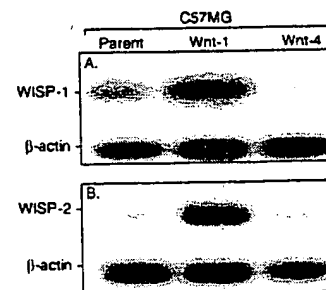


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (A) and *WISP-2* (B) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)<sup>+</sup> RNA (2  $\mu$ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human  $\beta$ -actin probe.

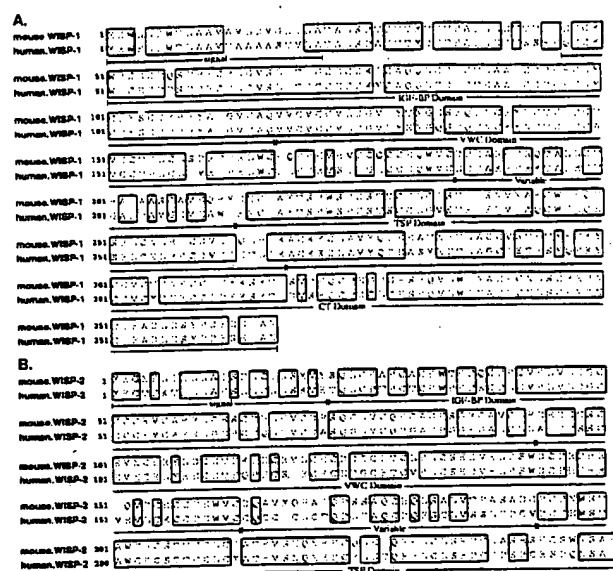


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

**Identification of *WISP-3*.** To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- $\beta$  (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

*WISP* proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

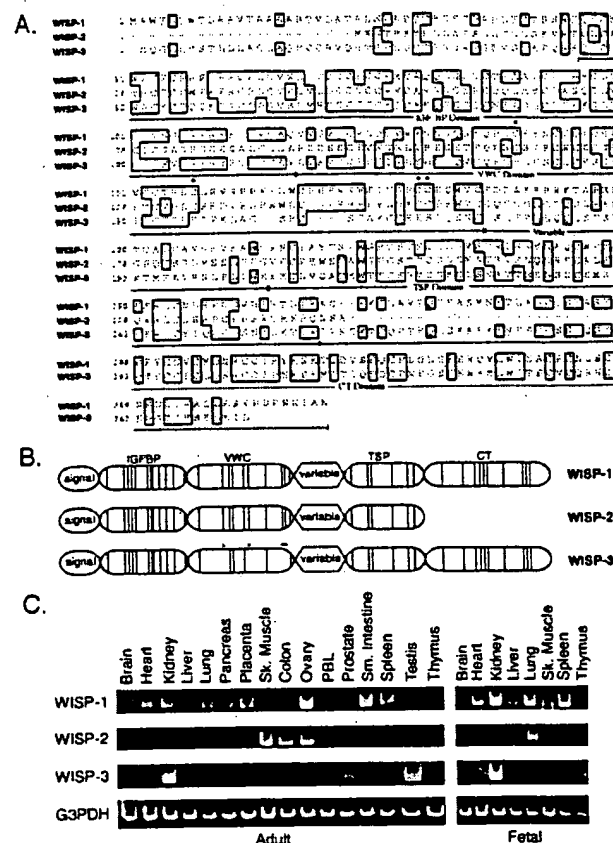


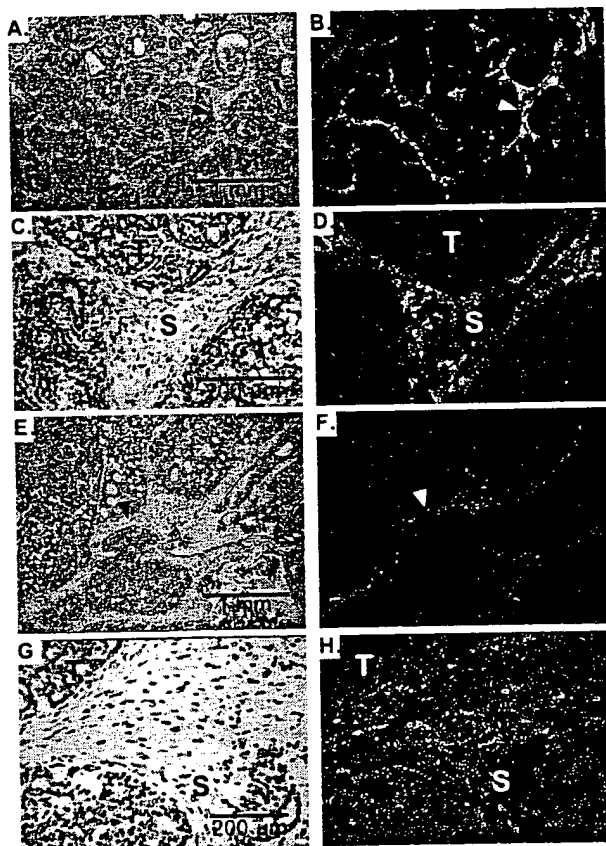
FIG. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSXCSxxCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

**Expression of *WISP* mRNA in Human Tissues.** Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

**In Situ Localization of *WISP-1* and *WISP-2*.** Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas



**FIG. 4.** (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells were negative (G and H).

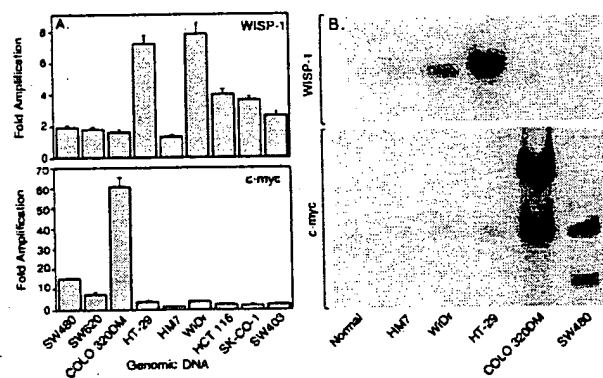
the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

**Chromosome Localization of the *WISP* Genes.** The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

**Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors.** Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ( $P < 0.001$  for each). The copy number for *WISP-3* was indistinguishable from one ( $P = 0.166$ ). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ( $P < 0.001$ ).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were



**FIG. 5.** Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10  $\mu$ g) digested with *Eco*RI (*WISP-1*) or *Xba*I (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

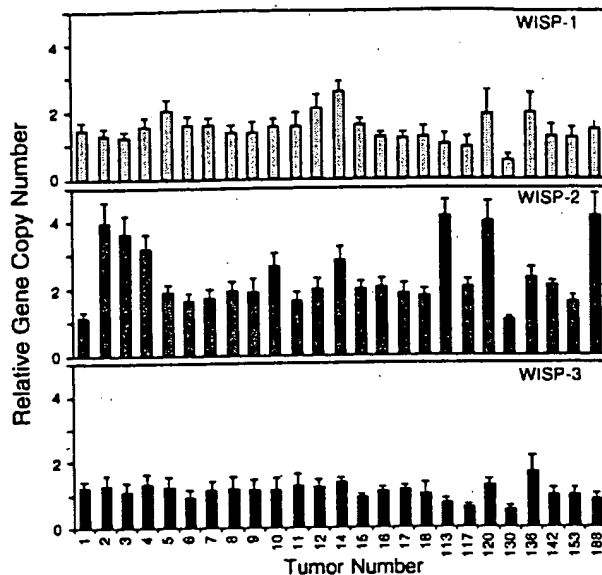


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means  $\pm$  SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

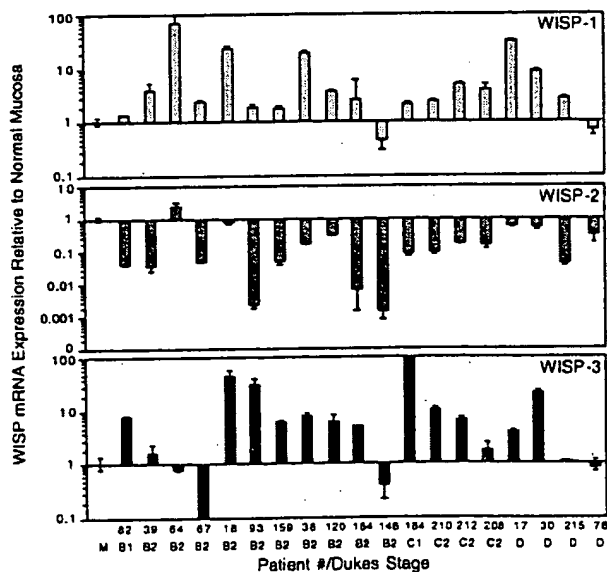


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means  $\pm$  SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

## DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e.,  $\beta$ -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through  $\beta$ -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- $\beta$ , platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin  $\alpha_v\beta_3$  serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- $\beta$ 1, which is the stimulus for stromal proliferation (34). TGF- $\beta$ 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of *WISP-1* and *WISP-2* in the stromal cells of breast tumors supports this paracrine model.

An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP-3* RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human *WISP-2* was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of *WISP-2*, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which *WISP-2* RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of *WISP-2* in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the *WISP* genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and  $\beta$ -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic  $\beta$ -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the *WISPs*. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of *WISPs* as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the *WISPs* in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

- Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286–3305.
- Dale, T. C. (1998) *Biochem. J.* 329, 209–223.
- Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99–109.
- van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233–240.
- Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619–625.
- Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* 10, 182–187.
- Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391–399.
- Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* 18, 1248–1256.
- Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046–3050.
- He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509–1512.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025–6030.
- Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001–1009.
- Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* 14, 6278–6286.
- Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349–1358.
- Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289–296.
- Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* 18, 6131–6141.
- Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171–179.
- Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* 16, 1326–1334.
- Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355–6360.
- Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729–2732.
- Bork, P. (1993) *FEBS Lett.* 327, 125–130.
- Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981–12986.
- Joliet, V., Martinerie, C., Dambrine, G., Plassiat, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* 12, 10–21.
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514–19527.
- Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852–2855.
- Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* 113, 195–205.
- Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529–2534.
- Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell. Genet.* 57, 109–111.
- Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88–94.
- Garte, S. J. (1993) *Crit. Rev. Oncog.* 4, 435–449.
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268–1272.
- Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269–291.
- Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090–3096.
- Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153–161.
- Wernert, N. (1997) *Virchows Arch.* 430, 433–443.
- Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., *et al.* (1994) *Cancer Res.* 54, 4257–4260.
- Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187–194.
- Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., *et al.* (1998) *EMBO J.* 17, 3052–3065.
- Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787–1790.
- Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 169–176.



methods. Peptides AENK or AEQK were dissolved in water, made isotonic with NaCl and diluted into RPMI growth medium. T-cell-proliferation assays were done essentially as described<sup>20,21</sup>. Briefly, after antigen pulsing ( $30 \mu\text{g ml}^{-1}$  TTCF) with tetrapeptides ( $1-2 \text{ mg ml}^{-1}$ ), PBMCs or EBV-B cells were washed in PBS and fixed for 45 s in 0.05% glutaraldehyde. Glycine was added to a final concentration of 0.1 M and the cells were washed five times in RPMI 1640 medium containing 1% FCS before co-culture with T-cell clones in round-bottom 96-well microtitre plates. After 48 h, the cultures were pulsed with  $1 \mu\text{Ci}$  of  $^3\text{H}$ -thymidine and harvested for scintillation counting 16 h later. Predigestion of native TTCF was done by incubating  $200 \mu\text{g}$  TTCF with  $0.25 \mu\text{g}$  pig kidney legumain in  $500 \mu\text{l}$  50 mM citrate buffer, pH 5.5, for 1 h at  $37^\circ\text{C}$ . **Glycopeptide digestions.** The peptides HIDNEEDI, HIDN(N-glucosamine) EEDI and HIDNESDI, which are based on the TTCF sequence, and QQQLFGSNVTDCSGNFCLFR(KKK), which is based on human transferrin, were obtained by custom synthesis. The three C-terminal lysine residues were added to the natural sequence to aid solubility. The transferrin glycopeptide QQQLFGSNVTDCSGNFCLFR was prepared by tryptic (Promega) digestion of 5 mg reduced, carboxy-methylated human transferrin followed by concanavalin A chromatography<sup>11</sup>. Glycopeptides corresponding to residues 622-642 and 421-452 were isolated by reverse-phase HPLC and identified by mass spectrometry and N-terminal sequencing. The lyophilized transferrin-derived peptides were redissolved in 50 mM sodium acetate, pH 5.5, 10 mM dithiothreitol, 20% methanol. Digestions were performed for 3 h at  $30^\circ\text{C}$  with  $5-50 \text{ mU ml}^{-1}$  pig kidney legumain or B-cell AEP. Products were analysed by HPLC or MALDI-TOF mass spectrometry using a matrix of  $10 \text{ mg ml}^{-1}$   $\alpha$ -cyanocinnamic acid in 50% acetonitrile/0.1% TFA and a PerSeptive Biosystems Elite STR mass spectrometer set to linear or reflector mode. Internal standardization was obtained with a matrix ion of 568.13 mass units.

Received 29 September; accepted 3 November 1998.

- Chen, J. M. et al. Cloning, isolation, and characterisation of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* 272, 8090-8098 (1997).
- Kembhavi, A. A., Buttle, D. J., Knight, C. G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* 303, 208-213 (1993).
- Dalton, J. P., Hla, J., Brindley, P. J. Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* 111, 575-580 (1995).
- Bennett, K. et al. Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol.* 22, 1519-1524 (1992).
- Riese, R. J. et al. Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357-366 (1996).
- Rodríguez, G. M. & Diment, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* 149, 2894-2898 (1992).
- Hewitt, E. W. et al. Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin: implications for T cell epitope generation. *J. Immunol.* 159, 4693-4699 (1997).
- Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* 15, 821-850 (1997).
- Chapman, H. A. Endosomal proteases and MHC class II function. *Curr. Opin. Immunol.* 10, 93-102 (1998).
- Fineschi, B. & Miller, J. Endosomal proteases and antigen processing. *Trends Biochem. Sci.* 22, 377-382 (1997).
- Lu, J. & van Halbeek, H. Complete  $^1\text{H}$  and  $^{13}\text{C}$  resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin. *Carbohydr. Res.* 296, 1-21 (1996).
- Fearon, D. T. & Locksley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* 272, 50-54 (1996).
- Medzhitov, R. & Janeway, C. A. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91, 295-298 (1997).
- Wyatt, R. et al. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 705-711 (1998).
- Botarelli, P. et al. N-glycosylation of HIV gp120 may constrain recognition by T lymphocytes. *J. Immunol.* 147, 3128-3132 (1991).
- Davidson, H. W., West, M. A. & Watts, C. Endocytosis, intracellular trafficking, and processing of membrane IgG and monovalent antigen/membrane IgG complexes in B lymphocytes. *J. Immunol.* 144, 4101-4109 (1990).
- Barrett, A. J. & Kirschke, H. Cathepsin B, cathepsin H and cathepsin L. *Methods Enzymol.* 80, 535-559 (1981).
- Makoff, A. J., Ballantine, S. P., Smallwood, A. E. & Fairweather, N. F. Expression of tetanus toxin fragment C in *E. coli*: its purification and potential use as a vaccine. *Biotechnology* 7, 1043-1046 (1989).
- Lane, D. P. & Harlow, E. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
- Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* 314, 537-539 (1985).
- Pond, L. & Watts, C. Characterization of transport of newly assembled, T cell-stimulatory MHC class II-peptide complexes from MHC class II compartments to the cell surface. *J. Immunol.* 159, 543-553 (1997).

**Acknowledgements.** We thank M. Ferguson for helpful discussions and advice; E. Smythe and L. Grayson for advice and technical assistance; B. Spruce, A. Knight and the BTS (Ninewells Hospital) for help with blood monocyte preparation; and our colleagues for many helpful comments on the manuscript. This work was supported by the Wellcome Trust and by an EMBO Long-term fellowship to B. M.

Correspondence and requests for materials should be addressed to C.W. (e-mail: c.watts@dundee.ac.uk).

## Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer

Robert M. Pitti<sup>†</sup>, Scot A. Marsters<sup>††</sup>, David A. Lawrence<sup>†</sup>, Margaret Roy<sup>\*</sup>, Frank C. Kischkel<sup>\*</sup>, Patrick Dowd<sup>\*</sup>, Arthur Huang<sup>\*</sup>, Christopher J. Donahue<sup>\*</sup>, Steven W. Sherwood<sup>\*</sup>, Daryl T. Baldwin<sup>\*</sup>, Paul J. Godowski<sup>\*</sup>, William I. Wood<sup>\*</sup>, Austin L. Gurney<sup>\*</sup>, Kenneth J. Hillan<sup>\*</sup>, Robert L. Cohen<sup>\*</sup>, Audrey D. Goddard<sup>\*</sup>, David Botstein<sup>†</sup> & Avi Ashkenazi<sup>\*</sup>

<sup>\*</sup> Departments of Molecular Oncology, Molecular Biology, and Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA

<sup>†</sup> Department of Genetics, Stanford University, Stanford, California 94305, USA

<sup>††</sup> These authors contributed equally to this work

Fas ligand (FasL) is produced by activated T cells and natural killer cells and it induces apoptosis (programmed cell death) in target cells through the death receptor Fas/Apo1/CD95 (ref. 1). One important role of FasL and Fas is to mediate immune-cytotoxic killing of cells that are potentially harmful to the organism, such as virus-infected or tumour cells<sup>1</sup>. Here we report the discovery of a soluble decoy receptor, termed decoy receptor 3 (DcR3), that binds to FasL and inhibits FasL-induced apoptosis. The DcR3 gene was amplified in about half of 35 primary lung and colon tumours studied, and DcR3 messenger RNA was expressed in malignant tissue. Thus, certain tumours may escape FasL-dependent immune-cytotoxic attack by expressing a decoy receptor that blocks FasL.

By searching expressed sequence tag (EST) databases, we identified a set of related ESTs that showed homology to the tumour necrosis factor (TNF) receptor (TNFR) gene superfamily<sup>2</sup>. Using the overlapping sequence, we isolated a previously unknown full-length complementary DNA from human fetal lung. We named the protein encoded by this cDNA decoy receptor 3 (DcR3). The cDNA encodes a 300-amino-acid polypeptide that resembles members of the TNFR family (Fig. 1a): the amino terminus contains a leader sequence, which is followed by four tandem cysteine-rich domains (CRDs). Like one other TNFR homologue, osteoprotegerin (OPG)<sup>3</sup>, DcR3 lacks an apparent transmembrane sequence, which indicates that it may be a secreted, rather than a membrane-associated, molecule. We expressed a recombinant, histidine-tagged form of DcR3 in mammalian cells; DcR3 was secreted into the cell culture medium, and migrated on polyacrylamide gels as a protein of relative molecular mass 35,000 (data not shown). DcR3 shares sequence identity in particular with OPG (31%) and TNFR2 (29%), and has relatively less homology with Fas (17%). All of the cysteines in the four CRDs of DcR3 and OPG are conserved; however, the carboxy-terminal portion of DcR3 is 101 residues shorter.

We analysed expression of DcR3 mRNA in human tissues by northern blotting (Fig. 1b). We detected a predominant 1.2-kilobase transcript in fetal lung, brain, and liver, and in adult spleen, colon and lung. In addition, we observed relatively high DcR3 mRNA expression in the human colon carcinoma cell line SW480.

To investigate potential ligand interactions of DcR3, we generated a recombinant, Fc-tagged DcR3 protein. We tested binding of DcR3-Fc to human 293 cells transfected with individual TNF-family ligands, which are expressed as type 2 transmembrane proteins (these transmembrane proteins have their N termini in the cytosol). DcR3-Fc showed a significant increase in binding to cells transfected with FasL<sup>4</sup> (Fig. 2a), but not to cells transfected with TNF<sup>5</sup>, Apo2L/TRAIL<sup>6,7</sup>, Apo3L/TWEAK<sup>8,9</sup>, or OPGL/TRANCE/



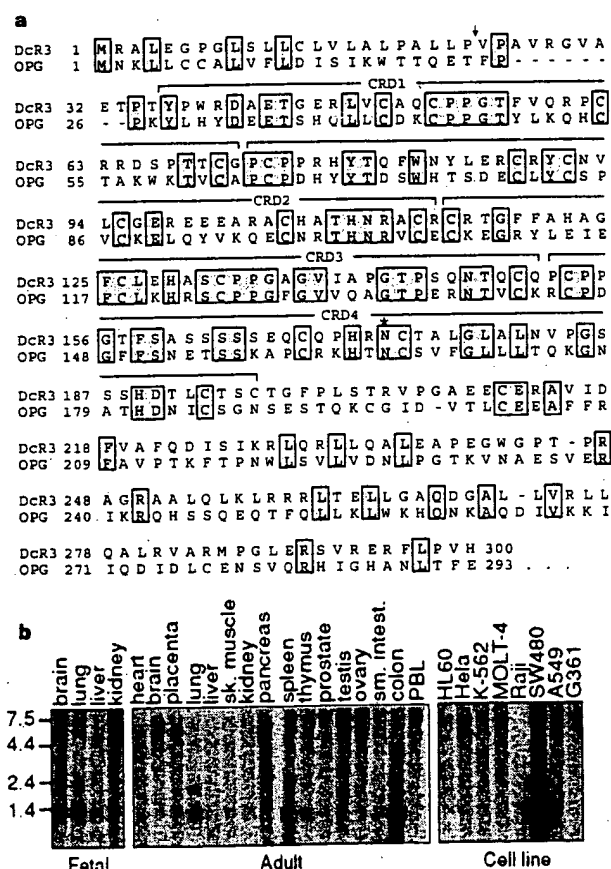
RANKL<sup>10-12</sup> (data not shown). DcR3-Fc immunoprecipitated shed FasL from FasL-transfected 293 cells (Fig. 2b) and purified soluble FasL (Fig. 2c), as did the Fc-tagged ectodomain of Fas but not TNFR1. Gel-filtration chromatography showed that DcR3-Fc and soluble FasL formed a stable complex (Fig. 2d). Equilibrium analysis indicated that DcR3-Fc and Fas-Fc bound to soluble FasL with a comparable affinity ( $K_d = 0.8 \pm 0.2$  and  $1.1 \pm 0.1$  nM, respectively; Fig. 2e), and that DcR3-Fc could block nearly all of the binding of soluble FasL to Fas-Fc (Fig. 2e, inset). Thus, DcR3 competes with Fas for binding to FasL.

To determine whether binding of DcR3 inhibits FasL activity, we tested the effect of DcR3-Fc on apoptosis induction by soluble FasL in Jurkat T leukaemia cells, which express Fas (Fig. 3a). DcR3-Fc and Fas-Fc blocked soluble-FasL-induced apoptosis in a similar dose-dependent manner, with half-maximal inhibition at  $\sim 0.1 \mu\text{g ml}^{-1}$ . Time-course analysis showed that the inhibition did not merely delay cell death, but rather persisted for at least 24 hours (Fig. 3b). We also tested the effect of DcR3-Fc on activation-induced cell death (AICD) of mature T lymphocytes, a FasL-dependent process<sup>1</sup>. Consistent with previous results<sup>13</sup>, activation of interleukin-2-stimulated CD4-positive T cells with anti-CD3 antibody increased the level of apoptosis twofold, and Fas-Fc blocked this effect substantially (Fig. 3c); DcR3-Fc blocked the

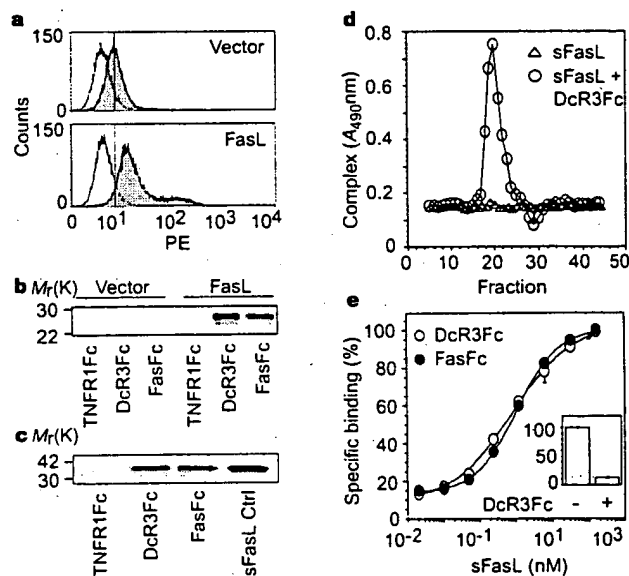
induction of apoptosis to a similar extent. Thus, DcR3 binding blocks apoptosis induction by FasL.

FasL-induced apoptosis is important in elimination of virus-infected cells and cancer cells by natural killer cells and cytotoxic T lymphocytes; an alternative mechanism involves perforin and granzymes<sup>1,14-16</sup>. Peripheral blood natural killer cells triggered marked cell death in Jurkat T leukaemia cells (Fig. 3d); DcR3-Fc and Fas-Fc each reduced killing of target cells from  $\sim 65\%$  to  $\sim 30\%$ , with half-maximal inhibition at  $\sim 1 \mu\text{g ml}^{-1}$ ; the residual killing was probably mediated by the perforin/granzyme pathway. Thus, DcR3 binding blocks FasL-dependent natural killer cell activity. Higher DcR3-Fc and Fas-Fc concentrations were required to block soluble FasL activity, which is consistent with the greater potency of membrane-associated FasL compared with soluble FasL<sup>17</sup>.

Given the role of immune-cytotoxic cells in elimination of tumour cells and the fact that DcR3 can act as an inhibitor of FasL, we proposed that DcR3 expression might contribute to the ability of some tumours to escape immune-cytotoxic attack. As genomic amplification frequently contributes to tumorigenesis, we investigated whether the DcR3 gene is amplified in cancer. We analysed DcR3 gene-copy number by quantitative polymerase chain



**Figure 1** Primary structure and expression of human DcR3. **a**, Alignment of the amino-acid sequences of DcR3 and of osteoprotegerin (OPG); the C-terminal 101 residues of OPG are not shown. The putative signal cleavage site (arrow), the cysteine-rich domains (CRD 1-4), and the N-linked glycosylation site (asterisk) are shown. **b**, Expression of DcR3 mRNA. Northern hybridization analysis was done using the DcR3 cDNA as a probe and blots of poly(A)<sup>+</sup> RNA (Clontech) from human fetal and adult tissues or cancer cell lines. PBL, peripheral blood lymphocyte.



**Figure 2** Interaction of DcR3 with FasL. **a**, 293 cells were transfected with pRK5 vector (top) or with pRK5 encoding full-length FasL (bottom), incubated with DcR3-Fc (solid line, shaded area), TNFR1-Fc (dotted line) or buffer control (dashed line) (the dashed and dotted lines overlap), and analysed for binding by FACS. Statistical analysis showed a significant difference ( $P < 0.001$ ) between the binding of DcR3-Fc to cells transfected with FasL or pRK5. PE, phycoerythrin-labelled cells. **b**, 293 cells were transfected as in **a** and metabolically labelled, and cell supernatants were immunoprecipitated with Fc-tagged TNFR1, DcR3 or Fas. **c**, Purified soluble FasL (sFasL) was immunoprecipitated with TNFR1-Fc, DcR3-Fc or Fas-Fc and visualized by immunoblot with anti-FasL antibody. sFasL was loaded directly for comparison in the right-hand lane. **d**, Flag-tagged sFasL was incubated with DcR3-Fc or with buffer and resolved by gel filtration; column fractions were analysed in an assay that detects complexes containing DcR3-Fc and sFasL-Flag. **e**, Equilibrium binding of DcR3-Fc or Fas-Fc to sFasL-Flag. Inset, competition of DcR3-Fc with Fas-Fc for binding to sFasL-Flag.

reaction (PCR)<sup>18</sup> in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBLs) of 10 healthy donors. Eight of 18 lung tumours and 9 of 17 colon tumours showed DcR3 gene amplification, ranging from 2- to 18-fold (Fig. 4a, b). To confirm this result, we analysed the colon tumour DNAs with three more, independent sets of DcR3-based PCR primers and probes; we observed nearly the same amplification (data not shown).

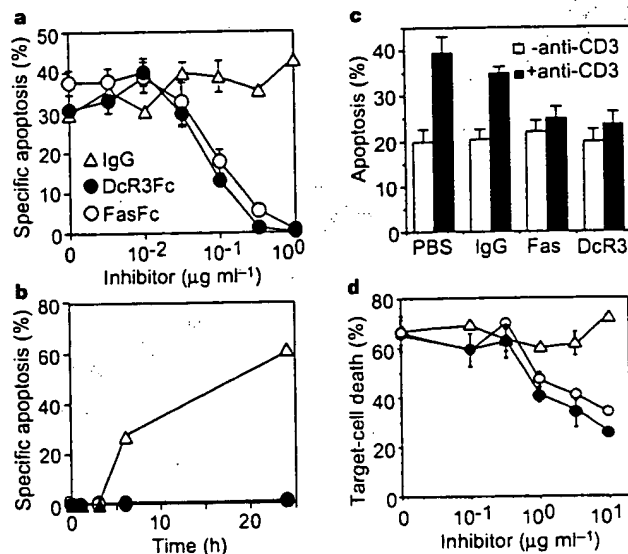
We then analysed DcR3 mRNA expression in primary tumour tissue sections by *in situ* hybridization. We detected DcR3 expression in 6 out of 15 lung tumours, 2 out of 2 colon tumours, 2 out of 5 breast tumours, and 1 out of 1 gastric tumour (data not shown). A section through a squamous-cell carcinoma of the lung is shown in Fig. 4c. DcR3 mRNA was localized to infiltrating malignant epithelium, but was essentially absent from adjacent stroma, indicating tumour-specific expression. Although the individual tumour specimens that we analysed for mRNA expression and gene amplification were different, the *in situ* hybridization results are consistent with the finding that the DcR3 gene is amplified frequently in tumours. SW480 colon carcinoma cells, which showed abundant DcR3 mRNA expression (Fig. 1b), also had marked DcR3 gene amplification, as shown by quantitative PCR (fourfold) and by Southern blot hybridization (fivefold) (data not shown).

If DcR3 amplification in cancer is functionally relevant, then DcR3 should be amplified more than neighbouring genomic regions that are not important for tumour survival. To test this,

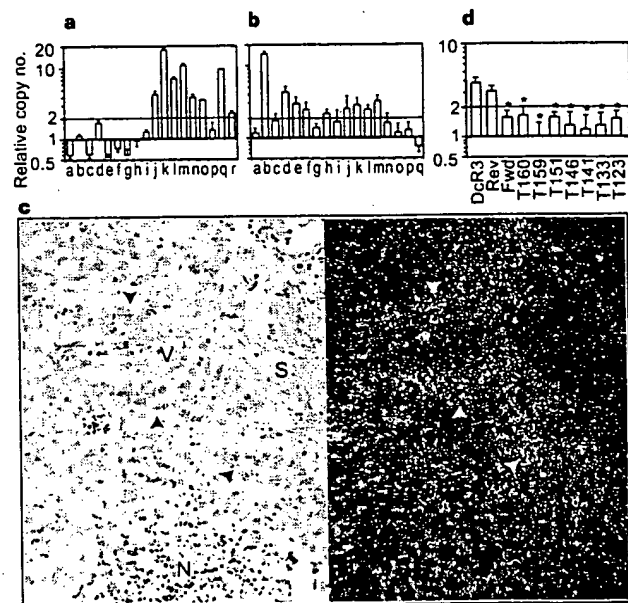
we mapped the human DcR3 gene by radiation-hybrid analysis; DcR3 showed linkage to marker AFM218xe7 (T160), which maps to chromosome position 20q13. Next, we isolated from a bacterial artificial chromosome (BAC) library a human genomic clone that carries DcR3, and sequenced the ends of the clone's insert. We then determined, from the nine colon tumours that showed twofold or greater amplification of DcR3, the copy number of the DcR3-flanking sequences (reverse and forward) from the BAC, and of seven genomic markers that span chromosome 20 (Fig. 4d). The DcR3-linked reverse marker showed an average amplification of roughly threefold, slightly less than the approximately fourfold amplification of DcR3; the other markers showed little or no amplification. These data indicate that DcR3 may be at the 'epi-centre' of a distal chromosome 20 region that is amplified in colon cancer, consistent with the possibility that DcR3 amplification promotes tumour survival.

Our results show that DcR3 binds specifically to FasL and inhibits FasL activity. We did not detect DcR3 binding to several other TNF-ligand-family members; however, this does not rule out the possibility that DcR3 interacts with other ligands, as do some other TNFR family members, including OPG<sup>2,19</sup>.

FasL is important in regulating the immune response; however, little is known about how FasL function is controlled. One mechanism involves the molecule cFLIP, which modulates apoptosis signalling downstream of Fas<sup>20</sup>. A second mechanism involves proteolytic shedding of FasL from the cell surface<sup>17</sup>. DcR3 competes with Fas for



**Figure 3** Inhibition of FasL activity by DcR3. **a**, Human Jurkat T leukaemia cells were incubated with Flag-tagged soluble FasL (sFasL; 5 ng ml<sup>-1</sup>) oligomerized with anti-Flag antibody (0.1 μg ml<sup>-1</sup>) in the presence of the proposed inhibitors DcR3-Fc, Fas-Fc or human IgG1 and assayed for apoptosis (mean ± s.e.m. of triplicates). **b**, Jurkat cells were incubated with sFasL-Flag plus anti-Flag antibody as in **a**, in presence of 1 μg ml<sup>-1</sup> DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and apoptosis was determined at the indicated time points. **c**, Peripheral blood natural killer cells were stimulated with PHA and interleukin-2, followed by control (white bars) or anti-CD3 antibody (filled bars), together with phosphate-buffered saline (PBS), human IgG1, Fas-Fc, or DcR3-Fc (10 μg ml<sup>-1</sup>). After 16 h, apoptosis of CD4<sup>+</sup> cells was determined (mean ± s.e.m. of results from five donors). **d**, Peripheral blood natural killer cells were incubated with <sup>51</sup>Cr-labelled Jurkat cells in the presence of DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and target-cell death was determined by release of <sup>51</sup>Cr (mean ± s.d. for two donors, each in triplicate).



**Figure 4** Genomic amplification of DcR3 in tumours. **a**, Lung cancers, comprising eight adenocarcinomas (c, d, f, g, h, j, k, r), seven squamous-cell carcinomas (a, e, m, n, o, p, q), one non-small-cell carcinoma (b), one small-cell carcinoma (i), and one bronchial adenocarcinoma (l). The data are means ± s.d. of 2 experiments done in duplicate. **b**, Colon tumours, comprising 17 adenocarcinomas. Data are means ± s.e.m. of five experiments done in duplicate. **c**, *In situ* hybridization analysis of DcR3 mRNA expression in a squamous-cell carcinoma of the lung. A representative bright-field image (left) and the corresponding dark-field image (right) show DcR3 mRNA over infiltrating malignant epithelium (arrowheads). Adjacent non-malignant stroma (S), blood vessel (V) and necrotic tumour tissue (N) are also shown. **d**, Average amplification of DcR3 compared with amplification of neighbouring genomic regions (reverse and forward, Rev and Fwd), the DcR3-linked marker T160, and other chromosome-20 markers, in the nine colon tumours showing DcR3 amplification of twofold or more (b). Data are from two experiments done in duplicate. Asterisk indicates  $P < 0.01$  for a Student's *t*-test comparing each marker with DcR3.

FasL binding; hence, it may represent a third mechanism of extracellular regulation of FasL activity. A decoy receptor that modulates the function of the cytokine interleukin-1 has been described<sup>21</sup>. In addition, two decoy receptors that belong to the TNFR family, DcR1 and DcR2, regulate the FasL-related apoptosis-inducing molecule Apo2L<sup>22</sup>. Unlike DcR1 and DcR2, which are membrane-associated proteins, DcR3 is directly secreted into the extracellular space. One other secreted TNFR-family member is OPG<sup>3</sup>, which shares greater sequence homology with DcR3 (31%) than do DcR1 (17%) or DcR2 (19%); OPG functions as a third decoy for Apo2L<sup>19</sup>. Thus, DcR3 and OPG define a new subset of TNFR-family members that function as secreted decoys to modulate ligands that induce apoptosis. Pox viruses produce soluble TNFR homologues that neutralize specific TNF-family ligands, thereby modulating the antiviral immune response<sup>2</sup>. Our results indicate that a similar mechanism, namely, production of a soluble decoy receptor for FasL, may contribute to immune evasion by certain tumours. □

## Methods

**Isolation of DcR3 cDNA.** Several overlapping ESTs in GenBank (accession numbers AA025672, AA025673 and W67560) and in Lifeseq<sup>TM</sup> (Incyte Pharmaceuticals; accession numbers 1339238, 1533571, 1533650, 1542861, 1789372 and 2207027) showed similarity to members of the TNFR family. We screened human cDNA libraries by PCR with primers based on the region of EST consensus; fetal lung was positive for a product of the expected size. By hybridization to a PCR-generated probe based on the ESTs, one positive clone (DNA30942) was identified. When searching for potential alternatively spliced forms of DcR3 that might encode a transmembrane protein, we isolated 50 more clones; the coding regions of these clones were identical in size to that of the initial clone (data not shown).

**Fc-fusion proteins (immunoadhesins).** The entire DcR3 sequence, or the ectodomain of Fas or TNFR1, was fused to the hinge and Fc region of human IgG1, expressed in insect SF9 cells or in human 293 cells, and purified as described<sup>23</sup>.

**Fluorescence-activated cell sorting (FACS) analysis.** We transfected 293 cells using calcium phosphate or Effectene (Qiagen) with pRK5 vector or pRK5 encoding full-length human FasL\* (2 µg), together with pRK5 encoding CrmA (2 µg) to prevent cell death. After 16 h, the cells were incubated with biotinylated DcR3-Fc or TNFR1-Fc and then with phycoerythrin-conjugated streptavidin (GibcoBRL), and were assayed by FACS. The data were analysed by Kolmogorov-Smirnov statistical analysis. There was some detectable staining of vector-transfected cells by DcR3-Fc; as these cells express little FasL (data not shown), it is possible that DcR3 recognized some other factor that is expressed constitutively on 293 cells.

**Immunoprecipitation.** Human 293 cells were transfected as above, and metabolically labelled with [<sup>35</sup>S]cysteine and [<sup>35</sup>S]methionine (0.5 mCi; Amersham). After 16 h of culture in the presence of z-VAD-fmk (10 µM), the medium was immunoprecipitated with DcR3-Fc, Fas-Fc or TNFR1-Fc (5 µg), followed by protein A-Sepharose (Repligen). The precipitates were resolved by SDS-PAGE and visualized on a phosphorimager (Fuji BAS2000). Alternatively, purified, Flag-tagged soluble FasL (1 µg) (Alexis) was incubated with each Fc-fusion protein (1 µg), precipitated with protein A-Sepharose, resolved by SDS-PAGE and visualized by immunoblotting with rabbit anti-FasL antibody (Oncogene Research).

**Analysis of complex formation.** Flag-tagged soluble FasL (25 µg) was incubated with buffer or with DcR3-Fc (40 µg) for 1.5 h at 24 °C. The reaction was loaded onto a Superdex 200 HR 10/30 column (Pharmacia) and developed with PBS; 0.6-ml fractions were collected. The presence of DcR3-Fc-FasL complex in each fraction was analysed by placing 100 µl aliquots into microtitre wells pre-coated with anti-human IgG (Boehringer) to capture DcR3-Fc, followed by detection with biotinylated anti-Flag antibody Bio M2 (Kodak) and streptavidin-horseradish peroxidase (Amersham). Calibration of the column indicated an apparent relative molecular mass of the complex of 420K (data not shown), which is consistent with a stoichiometry of two DcR3-Fc homodimers to two soluble FasL homotrimers.

**Equilibrium binding analysis.** Microtitre wells were coated with anti-human

IgG, blocked with 2% BSA in PBS. DcR3-Fc or Fas-Fc was added, followed by serially diluted Flag-tagged soluble FasL. Bound ligand was detected with anti-Flag antibody as above. In the competition assay, Fas-Fc was immobilized as above, and the wells were blocked with excess IgG1 before addition of Flag-tagged soluble FasL plus DcR3-Fc.

**T-cell AICD.** CD3<sup>+</sup> lymphocytes were isolated from peripheral blood of individual donors using anti-CD3 magnetic beads (Miltenyi Biotech), stimulated with phytohaemagglutinin (PHA; 2 µg ml<sup>-1</sup>) for 24 h, and cultured in the presence of interleukin-2 (100 U ml<sup>-1</sup>) for 5 days. The cells were plated in wells coated with anti-CD3 antibody (Pharmingen) and analysed for apoptosis 16 h later by FACS analysis of annexin-V-binding of CD4<sup>+</sup> cells<sup>24</sup>.

**Natural killer cell activity.** Natural killer cells were isolated from peripheral blood of individual donors using anti-CD56 magnetic beads (Miltenyi Biotech), and incubated for 16 h with <sup>51</sup>Cr-loaded Jurkat cells at an effector-to-target ratio of 1:1 in the presence of DcR3-Fc, Fas-Fc or human IgG1. Target-cell death was determined by release of <sup>51</sup>Cr in effector-target co-cultures relative to release of <sup>51</sup>Cr by detergent lysis of equal numbers of Jurkat cells.

**Gene-amplification analysis.** Surgical specimens were provided by J. Kern (lung tumours) and P. Quirke (colon tumours). Genomic DNA was extracted (Qiagen) and the concentration was determined using Hoechst dye 33258 intercalation fluorometry. Amplification was determined by quantitative PCR<sup>18</sup> using a TaqMan instrument (ABI). The method was validated by comparison of PCR and Southern hybridization data for the Myc and HER-2 oncogenes (data not shown). Gene-specific primers and fluorogenic probes were designed on the basis of the sequence of DcR3 or of nearby regions identified on a BAC carrying the human DcR3 gene; alternatively, primers and probes were based on Stanford Human Genome Center marker AFM218xe7 (T160), which is linked to DcR3 (likelihood score = 5.4), SHGC-36268 (T159), the nearest available marker which maps to ~500 kilobases from T160, and five extra markers that span chromosome 20. The DcR3-specific primer sequences were 5'-CTTCTTCGCGCAGCTG-3' and 5'-ATCACGCCGCCAGCAG-3' and the fluorogenic probe sequence was 5'-(FAM)-ACACGATGCGTGCTCCAGCAG AAp-(TAMARA), where FAM is 5'-fluorescein phosphoramidite. Relative gene-copy numbers were derived using the formula 2<sup>(ΔCT)</sup>, where ΔCT is the difference in amplification cycles required to detect DcR3 in peripheral blood lymphocyte DNA compared to test DNA.

Received 24 September; accepted 6 November 1998.

1. Nagata, S. Apoptosis by death factor. *Cell* 88, 355-365 (1997).
2. Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* 76, 959-962 (1994).
3. Simonet, W. S. et al. Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 89, 309-319 (1997).
4. Suda, T., Takahashi, T., Golstein, P. & Nagata, S. Molecular cloning and expression of Fas ligand, a novel member of the TNF family. *Cell* 75, 1169-1178 (1993).
5. Pennica, D. et al. Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature* 312, 724-729 (1984).
6. Pitti, R. M. et al. Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor receptor family. *J. Biol. Chem.* 271, 12687-12692 (1996).
7. Wiley, S. R. et al. Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3, 673-682 (1995).
8. Marsters, S. A. et al. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
9. Chicheportiche, Y. et al. TWEAK, a new secreted ligand in the TNF family that weakly induces apoptosis. *J. Biol. Chem.* 272, 32401-32410 (1997).
10. Wong, B. R. et al. TRANCE is a novel ligand of the TNFR family that activates c-Jun-N-terminal kinase in T cells. *J. Biol. Chem.* 272, 25190-25194 (1997).
11. Anderson, D. M. et al. A homolog of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* 390, 175-179 (1997).
12. Lacey, D. L. et al. Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* 93, 165-176 (1998).
13. Dhein, J., Walczak, H., Baumler, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by Apo1/(Fas/CD95). *Nature* 373, 438-441 (1995).
14. Arase, H., Arase, N. & Saito, T. Fas-mediated cytotoxicity by freshly isolated natural killer cells. *J. Exp. Med.* 181, 1235-1238 (1995).
15. Medvedev, A. E. et al. Regulation of Fas and Fas ligand expression in NK cells by cytokines and the involvement of Fas ligand in NK/LAK cell-mediated cytotoxicity. *Cytokine* 9, 394-404 (1997).
16. Moretta, A. Mechanisms in cell-mediated cytotoxicity. *Cell* 90, 13-18 (1997).
17. Tanaka, M., Itai, T., Adachi, M. & Nagata, S. Downregulation of Fas ligand by shedding. *Nature Med.* 4, 31-36 (1998).
18. Gelmini, S. et al. Quantitative PCR-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. *Clin. Chem.* 43, 752-758 (1997).
19. Emery, I. G. et al. Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* 273, 14363-14367 (1998).
20. Wallach, D. Placing death under control. *Nature* 388, 123-125 (1997).
21. Colotta, F. et al. Interleukin-1 type II receptor: a danger target for IL-1 that is regulated by IL-4. *Science* 261, 472-475 (1993).

22. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* 281, 1305–1308 (1998).
23. Ashkenazi, A. & Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* 9, 195–200 (1997).
24. Masters, S. et al. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 750–752 (1996).

Acknowledgements. We thank C. Clark, D. Pennica and V. Dixit for comments, and J. Kern and P. Quirke for tumour specimens.

Correspondence and requests for materials should be addressed to A.A. (e-mail: aa@gene.com). The GenBank accession number for the DcR3 cDNA sequence is AF104419.

## Crystal structure of the ATP-binding subunit of an ABC transporter

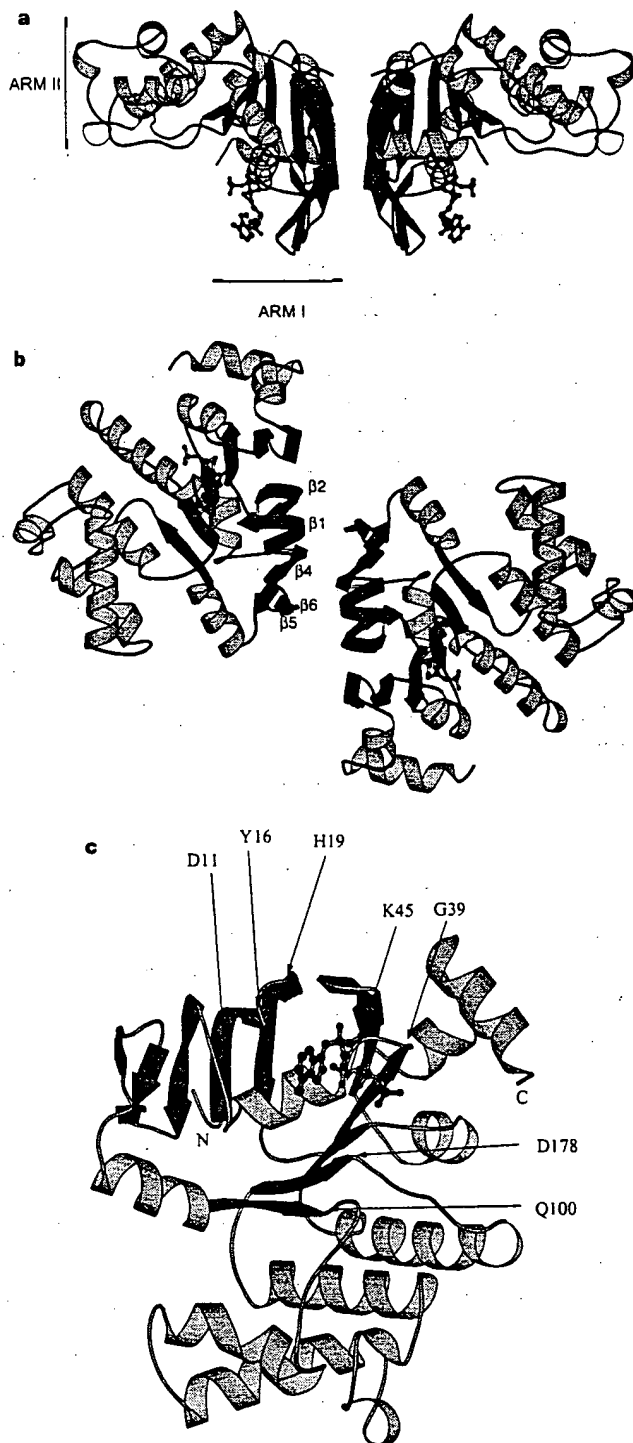
Li-Wei Hung\*, Iris Xiaoyan Wang†, Kishiko Nikaido†, Pei-Qi Liu†, Giovanna Ferro-Luzzi Ames† & Sung-Hou Kim\*‡

\* E. O. Lawrence Berkeley National Laboratory, † Department of Molecular and Cell Biology, and ‡ Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA

ABC transporters (also known as traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes<sup>1</sup>. The recently completed *Escherichia coli* genome sequence revealed that the largest family of paralogous *E. coli* proteins is composed of ABC transporters<sup>2</sup>. Many eukaryotic proteins of medical significance belong to this family, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the P-glycoprotein (or multidrug-resistance protein) and the heterodimeric transporter associated with antigen processing (Tap1–Tap2). Here we report the crystal structure at 1.5 Å resolution of HisP, the ATP-binding subunit of the histidine permease, which is an ABC transporter from *Salmonella typhimurium*. We correlate the details of this structure with the biochemical, genetic and biophysical properties of the wild-type and several mutant HisP proteins. The structure provides a basis for understanding properties of ABC transporters and of defective CFTR proteins.

ABC transporters contain four structural domains: two nucleotide-binding domains (NBDs), which are highly conserved throughout the family, and two transmembrane domains<sup>1</sup>. In prokaryotes these domains are often separate subunits which are assembled into a membrane-bound complex; in eukaryotes the domains are generally fused into a single polypeptide chain. The periplasmic histidine permease of *S. typhimurium* and *E. coli*<sup>3–6</sup> is a well-characterized ABC transporter that is a good model for this superfamily. It consists of a membrane-bound complex, HisQMP<sub>2</sub>, which comprises integral membrane subunits, HisQ and HisM, and two copies of HisP, the ATP-binding subunit. HisP, which has properties intermediate between those of integral and peripheral membrane proteins<sup>7</sup>, is accessible from both sides of the membrane, presumably by its interaction with HisQ and HisM<sup>6</sup>. The two HisP subunits form a dimer, as shown by their cooperativity in ATP hydrolysis<sup>5</sup>, the requirement for both subunits to be present for activity<sup>8</sup>, and the formation of a HisP dimer upon chemical cross-linking. Soluble HisP also forms a dimer<sup>3</sup>. HisP has been purified and characterized in an active soluble form<sup>3</sup> which can be reconstituted into a fully active membrane-bound complex<sup>8</sup>.

The overall shape of the crystal structure of the HisP monomer is that of an 'L' with two thick arms (arm I and arm II); the ATP-binding pocket is near the end of arm I (Fig. 1). A six-stranded  $\beta$ -sheet ( $\beta 3$  and  $\beta 8$ – $\beta 12$ ) spans both arms of the L, with a domain of a  $\alpha$ - plus  $\beta$ -type structure ( $\beta 1$ ,  $\beta 2$ ,  $\beta 4$ – $\beta 7$ ,  $\alpha 1$  and  $\alpha 2$ ) on one side (within arm I) and a domain of mostly  $\alpha$ -helices ( $\alpha 3$ – $\alpha 9$ ) on the



**Figure 1** Crystal structure of HisP. **a**, View of the dimer along an axis perpendicular to its two-fold axis. The top and bottom of the dimer are suggested to face towards the periplasmic and cytoplasmic sides, respectively (see text). The thickness of arm II is about 25 Å, comparable to that of membrane.  $\alpha$ -Helices are shown in orange and  $\beta$ -sheets in green. **b**, View along the two-fold axis of the HisP dimer, showing the relative displacement of the monomers not apparent in **a**. The  $\beta$ -strands at the dimer interface are labelled. **c**, View of one monomer from the bottom of arm I, as shown in **a**, towards arm II, showing the ATP-binding pocket. **a**–**c**, The protein and the bound ATP are in 'ribbon' and 'ball-and-stick' representations, respectively. Key residues discussed in the text are indicated in **c**. These figures were prepared with MOLSCRIPT<sup>29</sup>. N, amino terminus; C, C terminus.

## NOVEL APPROACH TO QUANTITATIVE POLYMERASE CHAIN REACTION USING REAL-TIME DETECTION: APPLICATION TO THE DETECTION OF GENE AMPLIFICATION IN BREAST CANCER

Ivan BIÈCHE<sup>1,2</sup>, Martine OLIVI<sup>1</sup>, Marie-Hélène CHAMPÈME<sup>2</sup>, Dominique VIDAUD<sup>1</sup>, Rosette LIDEREAU<sup>2</sup> and Michel VIDAUD<sup>1\*</sup>

<sup>1</sup>Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, Paris, France

<sup>2</sup>Laboratoire d'Oncogénétique, Centre René Huguenin, St-Cloud, France

Gene amplification is a common event in the progression of human cancers, and amplified oncogenes have been shown to have diagnostic, prognostic and therapeutic relevance. A kinetic quantitative polymerase-chain-reaction (PCR) method, based on fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real-time, was used to quantify gene amplification in tumor DNA. Reactions are characterized by the point during cycling when PCR amplification is still in the exponential phase, rather than the amount of PCR product accumulated after a fixed number of cycles. None of the reaction components is limited during the exponential phase, meaning that values are highly reproducible in reactions starting with the same copy number. This greatly improves the precision of DNA quantification. Moreover, real-time PCR does not require post-PCR sample handling, thereby preventing potential PCR-product carry-over contamination; it possesses a wide dynamic range of quantification and results in much faster and higher sample throughput. The real-time PCR method, was used to develop and validate a simple and rapid assay for the detection and quantification of the 3 most frequently amplified genes (*myc*, *ccnd1* and *erbB2*) in breast tumors. Extra copies of *myc*, *ccnd1* and *erbB2* were observed in 10, 23 and 15%, respectively, of 108 breast-tumor DNA; the largest observed numbers of gene copies were 4.6, 18.6 and 15.1, respectively. These results correlated well with those of Southern blotting. The use of this new semi-automated technique will make molecular analysis of human cancers simpler and more reliable, and should find broad applications in clinical and research settings. *Int. J. Cancer* 78:661–666, 1998.

© 1998 Wiley-Liss, Inc.

Gene amplification plays an important role in the pathogenesis of various solid tumors, including breast cancer, probably because over-expression of the amplified target genes confers a selective advantage. The first technique used to detect genomic amplification was cytogenetic analysis. Amplification of several chromosome regions, visualized either as extrachromosomal double minutes (dmns) or as integrated homogeneously staining regions (HSRs), are among the main visible cytogenetic abnormalities in breast tumors. Other techniques such as comparative genomic hybridization (CGH) (Kallioniemi *et al.*, 1994) have also been used in broad searches for regions of increased DNA copy numbers in tumor cells, and have revealed some 20 amplified chromosome regions in breast tumors. Positional cloning efforts are underway to identify the critical gene(s) in each amplified region. To date, genes known to be amplified frequently in breast cancers include *myc* (8q24), *ccnd1* (11q13), and *erbB2* (17q12-q21) (for review, see Bièche and Lidereau, 1995).

Amplification of the *myc*, *ccnd1*, and *erbB2* proto-oncogenes should have clinical relevance in breast cancer, since independent studies have shown that these alterations can be used to identify sub-populations with a worse prognosis (Berns *et al.*, 1992; Schuurin *et al.*, 1992; Slamon *et al.*, 1987). Muss *et al.* (1994) suggested that these gene alterations may also be useful for the prediction and assessment of the efficacy of adjuvant chemotherapy and hormone therapy.

However, published results diverge both in terms of the frequency of these alterations and their clinical value. For instance, over 500 studies in 10 years have failed to resolve the controversy

surrounding the link suggested by Slamon *et al.* (1987) between *erbB2* amplification and disease progression. These discrepancies are partly due to the clinical, histological and ethnic heterogeneity of breast cancer, but technical considerations are also probably involved.

Specific genes (DNA) were initially quantified in tumor cells by means of blotting procedures such as Southern and slot blotting. These batch techniques require large amounts of DNA (5–10 µg/reaction) to yield reliable quantitative results. Furthermore, meticulous care is required at all stages of the procedures to generate blots of sufficient quality for reliable dosage analysis. Recently, PCR has proven to be a powerful tool for quantitative DNA analysis, especially with minimal starting quantities of tumor samples (small, early-stage tumors and formalin-fixed, paraffin-embedded tissues).

Quantitative PCR can be performed by evaluating the amount of product either after a given number of cycles (end-point quantitative PCR) or after a varying number of cycles during the exponential phase (kinetic quantitative PCR). In the first case, an internal standard distinct from the target molecule is required to ascertain PCR efficiency. The method is relatively easy but implies generating, quantifying and storing an internal standard for each gene studied. Nevertheless, it is the most frequently applied method to date.

One of the major advantages of the kinetic method is its rapidity in quantifying a new gene, since no internal standard is required (an external standard curve is sufficient). Moreover, the kinetic method has a wide dynamic range (at least 5 orders of magnitude), giving an accurate value for samples differing in their copy number. Unfortunately, the method is cumbersome and has therefore been rarely used. It involves aliquot sampling of each assay mix at regular intervals and quantifying, for each aliquot, the amplification product. Interest in the kinetic method has been stimulated by a novel approach using fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real time (Gibson *et al.*, 1996; Heid *et al.*, 1996). The TaqMan reaction is based on the 5' nuclease assay first described by Holland *et al.* (1991). The latter uses the 5' nuclease activity of Taq polymerase to cleave a specific fluorogenic oligonucleotide probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee *et al.*, 1993). One fluorescent dye, co-valently linked to the 5' end of the oligonucleotide, serves as a reporter [FAM (*i.e.*, 6-carboxy-fluorescein)] and its emission spectrum is quenched by a second fluorescent dye, TAMRA (*i.e.*, 6-carboxy-tetramethyl-rhodamine) attached to the 3' end. During the extension phase of the PCR

Grant sponsors: Association Pour la Recherche sur le Cancer and Ministère de l'Enseignement Supérieur et de la Recherche.

\*Correspondence to: Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, 4 Avenue de l'Observatoire, F-75006 Paris, France. Fax: (33)1-4407-1754. E-mail: mvidaud@teaser.fr

Received 2 May 1998; Revised 30 June 1998

cycle, the fluorescent hybridization probe is hydrolyzed by the 5'-3' nucleolytic activity of DNA polymerase. Nuclease degradation of the probe releases the quenching of FAM fluorescence emission, resulting in an increase in peak fluorescence emission. The fluorescence signal is normalized by dividing the emission intensity of the reporter dye (FAM) by the emission intensity of a reference dye (i.e., ROX, 6-carboxy-X-rhodamine) included in TaqMan buffer, to obtain a ratio defined as the Rn (normalized reporter) for a given reaction tube. The use of a sequence detector enables the fluorescence spectra of all 96 wells of the thermal cycler to be measured continuously during PCR amplification.

The real-time PCR method offers several advantages over other current quantitative PCR methods (Celi *et al.*, 1994): (i) the probe-based homogeneous assay provides a real-time method for detecting only specific amplification products, since specific hybridization of both the primers and the probe is necessary to generate a signal; (ii) the  $C_t$  (threshold cycle) value used for quantification is measured when PCR amplification is still in the log phase of PCR product accumulation. This is the main reason why  $C_t$  is a more reliable measure of the starting copy number than are end-point measurements, in which a slight difference in a limiting component can have a drastic effect on the amount of product; (iii) use of  $C_t$  values gives a wider dynamic range (at least 5 orders of magnitude), reducing the need for serial dilution; (iv) The real-time PCR method is run in a closed-tube system and requires no post-PCR sample handling, thus avoiding potential contamination; (v) the system is highly automated, since the instrument continuously measures fluorescence in all 96 wells of the thermal cycler during PCR amplification and the corresponding software processes, and analyzes the fluorescence data; (vi) the assay is rapid, as results are available just one minute after thermal cycling is complete; (vii) the sample throughput of the method is high, since 96 reactions can be analyzed in 2 hr.

Here, we applied this semi-automated procedure to determine the copy numbers of the 3 most frequently amplified genes in breast tumors (*myc*, *ccnd1* and *erbB2*), as well as 2 genes (*alb* and *app*) located in a chromosome region in which no genetic changes have been observed in breast tumors. The results for 108 breast tumors were compared with previous Southern-blot data for the same samples.

## MATERIAL AND METHODS

### Tumor and blood samples

Samples were obtained from 108 primary breast tumors removed surgically from patients at the Centre René Huguénin; none of the patients had undergone radiotherapy or chemotherapy. Immediately after surgery, the tumor samples were placed in liquid nitrogen until extraction of high-molecular-weight DNA. Patients were included in this study if the tumor sample used for DNA preparation contained more than 60% of tumor cells (histological analysis). A blood sample was also taken from 18 of the same patients.

DNA was extracted from tumor tissue and blood leukocytes according to standard methods.

### Real-time PCR

**Theoretical basis.** Reactions are characterized by the point during cycling when amplification of the PCR product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the genomic DNA target, the earlier a significant increase in fluorescence is observed. The parameter  $C_t$  (threshold cycle) is defined as the fractional cycle number at which the fluorescence generated by cleavage of the probe passes a fixed threshold above baseline. The target gene copy number in unknown samples is quantified by measuring  $C_t$  and by using a standard curve to determine the starting copy number. The precise amount of genomic DNA (based on optical density) and its quality (i.e., lack

of extensive degradation) are both difficult to assess. We therefore also quantified a control gene (*alb*) mapping to chromosome region 4q11-q13, in which no genetic alterations have been found in breast-tumor DNA by means of CGH (Kallioniemi *et al.*, 1994).

Thus, the ratio of the copy number of the target gene to the copy number of the *alb* gene normalizes the amount and quality of genomic DNA. The ratio defining the level of amplification is termed "N", and is determined as follows:

$$N = \frac{\text{copy number of target gene (app, myc, ccnd1, erbB2)}}{\text{copy number of reference gene (alb)}}$$

**Primers, probes, reference human genomic DNA and PCR consumables.** Primers and probes were chosen with the assistance of the computer programs Oligo 4.0 (National Biosciences, Plymouth, MN), EuGene (Daniben Systems, Cincinnati, OH) and Primer Express (Perkin-Elmer Applied Biosystems, Foster City, CA).

Primers were purchased from DNAScience (Malvern, PA) and probes from Perkin-Elmer Applied Biosystems.

Nucleotide sequences for the oligonucleotide hybridization probes and primers are available on request.

The TaqMan PCR Core reagent kit, MicroAmp optical tubes, and MicroAmp caps were from Perkin-Elmer Applied Biosystems.

**Standard-curve construction.** The kinetic method requires a standard curve. The latter was constructed with serial dilutions of specific PCR products, according to Piatak *et al.* (1993). In practice, each specific PCR product was obtained by amplifying 20 ng of a standard human genomic DNA (Boehringer, Mannheim, Germany) with the same primer pairs as those used later for real-time quantitative PCR. The 5 PCR products were purified using MicroSpin S-400 HR columns (Pharmacia, Uppsala, Sweden) electrophoresed through an acrylamide gel and stained with ethidium bromide to check their quality. The PCR products were then quantified spectrophotometrically and pooled, and serially diluted 10-fold in mouse genomic DNA (Clontech, Palo Alto, CA) at a constant concentration of 2 ng/ $\mu$ l. The standard curve used for real-time quantitative PCR was based on serial dilutions of the pool of PCR products ranging from  $10^{-7}$  ( $10^5$  copies of each gene) to  $10^{-10}$  ( $10^2$  copies). This series of diluted PCR products was aliquoted and stored at  $-80^\circ\text{C}$  until use.

The standard curve was validated by analyzing 2 known quantities of calibrator human genomic DNA (20 ng and 50 ng).

**PCR amplification.** Amplification mixes (50  $\mu$ l) contained the sample DNA (around 20 ng, around 6600 copies of disomic genes),  $10\times$  TaqMan buffer (5  $\mu$ l), 200  $\mu$ M dATP, dCTP, dGTP, and 400  $\mu$ M dUTP, 5 mM  $\text{MgCl}_2$ , 1.25 units of AmpliTaq Gold, 0.5 units of AmpErase uracil N-glycosylase (UNG), 200 nM each primer and 100 nM probe. The thermal cycling conditions comprised 2 min at  $50^\circ\text{C}$  and 10 min at  $95^\circ\text{C}$ . Thermal cycling consisted of 40 cycles at  $95^\circ\text{C}$  for 15 s and  $65^\circ\text{C}$  for 1 min. Each assay included: a standard curve (from  $10^5$  to  $10^2$  copies) in duplicate, a no-template control, 20 ng and 50 ng of calibrator human genomic DNA (Boehringer) in triplicate, and about 20 ng of unknown genomic DNA in triplicate (26 samples can thus be analyzed on a 96-well microplate). All samples with a coefficient of variation (CV) higher than 10% were retested.

All reactions were performed in the ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems), which detects the signal from the fluorogenic probe during PCR.

**Equipment for real-time detection.** The 7700 system has a built-in thermal cycler and a laser directed via fiber optical cables to each of the 96 sample wells. A charge-coupled-device (CDD) camera collects the emission from each sample and the data are analyzed automatically. The software accompanying the 7700 system calculates  $C_t$  and determines the starting copy number in the samples.

**Determination of gene amplification.** Gene amplification was calculated as described above. Only samples with an N value higher than 2 were considered to be amplified.

### RESULTS

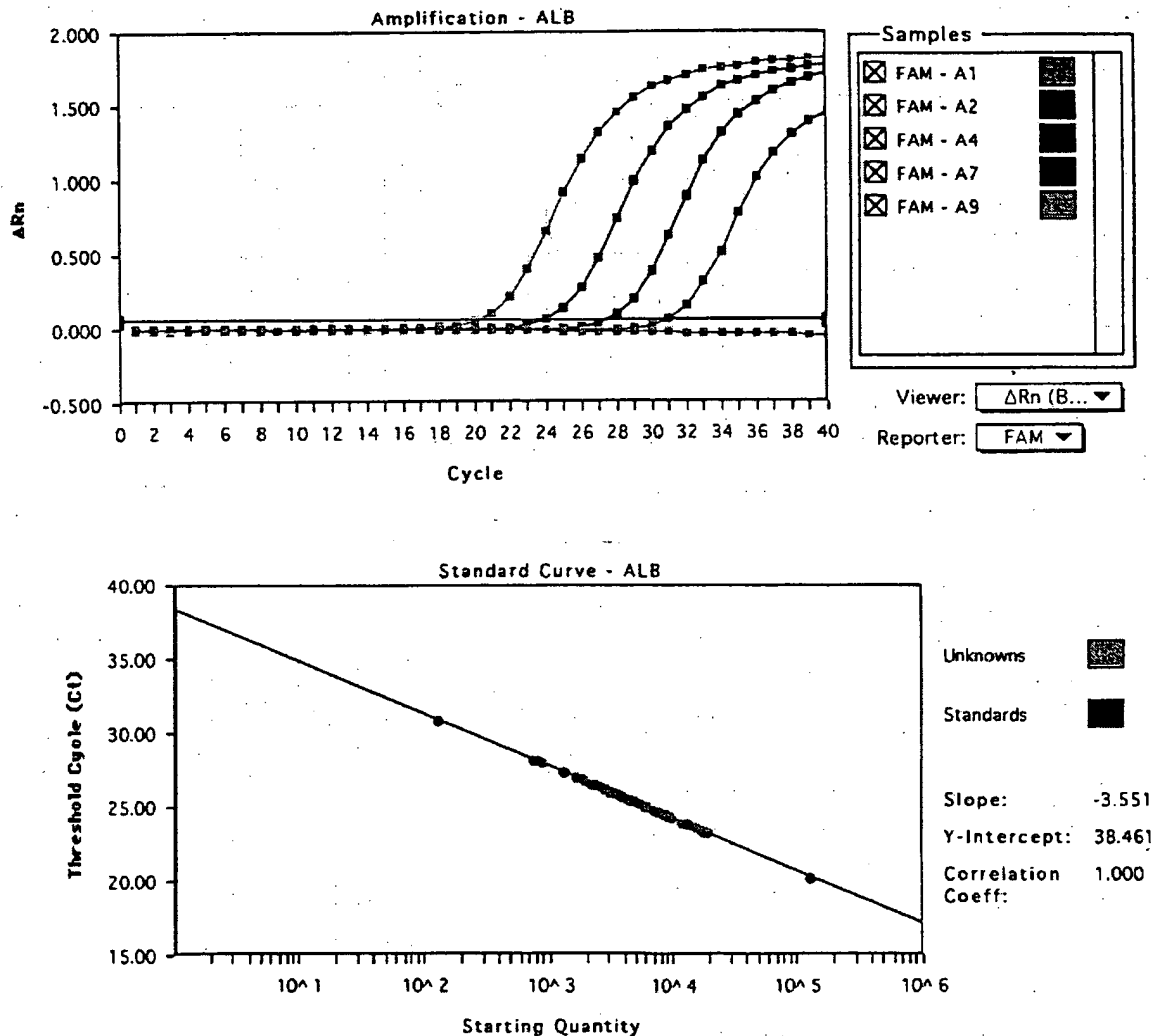
To validate the method, real-time PCR was performed on genomic DNA extracted from 108 primary breast tumors, and 18 normal leukocyte DNA samples from some of the same patients. The target genes were the *myc*, *ccnd1* and *erbB2* proto-oncogenes, and the  $\beta$ -amyloid precursor protein gene (*app*), which maps to a chromosome region (21q21.2) in which no genetic alterations have been found in breast tumors (Kallioniemi *et al.*, 1994). The reference disomic gene was the albumin gene (*alb*, chromosome 4q11-q13).

### Validation of the standard curve and dynamic range of real-time PCR

The standard curve was constructed from PCR products serially diluted in genomic mouse DNA at a constant concentration of 2 ng/ $\mu$ l. It should be noted that the 5 primer pairs chosen to analyze the 5 target genes do not amplify genomic mouse DNA (data not shown). Figure 1 shows the real-time PCR standard curve for the *alb* gene. The dynamic range was wide (at least 4 orders of magnitude), with samples containing as few as  $10^2$  copies or as many as  $10^5$  copies.

### Copy-number ratio of the 2 reference genes (*app* and *alb*)

The *app* to *alb* copy-number ratio was determined in 18 normal leukocyte DNA samples and all 108 primary breast-tumor DNA



**FIGURE 1** – Albumin (*alb*) gene dosage by real-time PCR. Top: Amplification plots for reactions with starting *alb* gene copy number ranging from  $10^5$  (A9),  $10^4$  (A7),  $10^3$  (A4) to  $10^2$  (A2) and a no-template control (A1). Cycle number is plotted vs. change in normalized reporter signal ( $\Delta Rn$ ). For each reaction tube, the fluorescence signal of the reporter dye (FAM) is divided by the fluorescence signal of the passive reference dye (ROX), to obtain a ratio defined as the normalized reporter signal ( $Rn$ ).  $\Delta Rn$  represents the normalized reporter signal ( $Rn$ ) minus the baseline signal established in the first 15 PCR cycles.  $\Delta Rn$  increases during PCR as *alb* PCR product copy number increases until the reaction reaches a plateau.  $C_t$  (threshold cycle) represents the fractional cycle number at which a significant increase in  $Rn$  above a baseline signal (horizontal black line) can first be detected. Two replicate plots were performed for each standard sample, but the data for only one are shown here. Bottom: Standard curve plotting log starting copy number vs.  $C_t$  (threshold cycle). The black dots represent the data for standard samples plotted in duplicate and the red dots the data for unknown genomic DNA samples plotted in triplicate. The standard curve shows 4 orders of linear dynamic range.



samples. We selected these 2 genes because they are located in 2 chromosome regions (*app*, 21q21.2; *alb*, 4q11-q13) in which no obvious genetic changes (including gains or losses) have been observed in breast cancers (Kallioniemi *et al.*, 1994). The ratio for the 18 normal leukocyte DNA samples fell between 0.7 and 1.3 (mean  $1.02 \pm 0.21$ ), and was similar for the 108 primary breast-tumor DNA samples (0.6 to 1.6, mean  $1.06 \pm 0.25$ ), confirming that *alb* and *app* are appropriate reference disomic genes for breast-tumor DNA. The low range of the ratios also confirmed that the nucleotide sequences chosen for the primers and probes were not polymorphic, as mismatches of their primers or probes with the subject's DNA would have resulted in differential amplification.

#### *myc*, *ccnd1* and *erbB2* gene dose in normal leukocyte DNA

To determine the cut-off point for gene amplification in breast-cancer tissue, 18 normal leukocyte DNA samples were tested for the gene dose (N), calculated as described in "Material and Methods". The N value of these samples ranged from 0.5 to 1.3 (mean  $0.84 \pm 0.22$ ) for *myc*; 0.7 to 1.6 (mean  $1.06 \pm 0.23$ ) for *ccnd1* and 0.6 to 1.3 (mean  $0.91 \pm 0.19$ ) for *erbB2*. Since N values for *myc*, *ccnd1* and *erbB2* in normal leukocyte DNA consistently fell between 0.5 and 1.6, values of 2 or more were considered to represent gene amplification in tumor DNA.

#### *myc*, *ccnd1* and *erbB2* gene dose in breast-tumor DNA

*myc*, *ccnd1* and *erbB2* gene copy numbers in the 108 primary breast tumors are reported in Table I. Extra copies of *ccnd1* were more frequent (23%, 25/108) than extra copies of *erbB2* (15%, 16/108) and *myc* (10%, 11/108), and ranged from 2 to 18.6 for *ccnd1*, 2 to 15.1 for *erbB2*, and only 2 to 4.6 for the *myc* gene. Figure 2 and Table II represent tumors in which the *ccnd1* gene was amplified 16-fold (T145), 6-fold (T133) and non-amplified (T118). The 3 genes were never found to be co-amplified in the same tumor. *erbB2* and *ccnd1* were co-amplified in only 3 cases, *myc* and *ccnd1* in 2 cases and *myc* and *erbB2* in 1 case. This favors the hypothesis that gene amplifications are independent events in breast cancer. Interestingly, 5 tumors showed a decrease of at least 50% in the *erbB2* copy number ( $N < 0.5$ ), suggesting that they bore deletions of the 17q21 region (the site of *erbB2*). No such decrease in copy number was observed with the other 2 proto-oncogenes.

#### Comparison of gene dose determined by real-time quantitative PCR and Southern-blot analysis

Southern-blot analysis of *myc*, *ccnd1* and *erbB2* amplifications had previously been done on the same 108 primary breast tumors. A perfect correlation between the results of real-time PCR and Southern blot was obtained for tumors with high copy numbers ( $N \geq 5$ ). However, there were cases (1 *myc*, 6 *ccnd1* and 4 *erbB2*) in which real-time PCR showed gene amplification whereas Southern-blot did not, but these were mainly cases with low extra copy numbers (N from 2 to 2.9).

#### DISCUSSION

The clinical applications of gene amplification assays are currently limited, but would certainly increase if a simple, standardized and rapid method were perfected. Gene amplification status has been studied mainly by means of Southern blotting, but this method is not sensitive enough to detect low-level gene amplification nor accurate enough to quantify the full range of amplification values. Southern blotting is also time-consuming, uses radioactive

reagents and requires relatively large amounts of high-quality genomic DNA, which means it cannot be used routinely in many laboratories. An amplification step is therefore required to determine the copy number of a given target gene from minimal quantities of tumor DNA (small early-stage tumors, cytopuncture specimens or formalin-fixed, paraffin-embedded tissues).

In this study, we validated a PCR method developed for the quantification of gene over-representation in tumors. The method, based on real-time analysis of PCR amplification, has several advantages over other PCR-based quantitative assays such as competitive quantitative PCR (Celi *et al.*, 1994). First, the real-time PCR method is performed in a closed-tube system, avoiding the risk of contamination by amplified products. Re-amplification of carryover PCR products in subsequent experiments can also be prevented by using the enzyme uracil N-glycosylase (UNG) (Longo *et al.*, 1990). The second advantage is the simplicity and rapidity of sample analysis, since no post-PCR manipulations are required. Our results show that the automated method is reliable. We found it possible to determine, in triplicate, the number of copies of a target gene in more than 100 tumors per day. Third, the system has a linear dynamic range of at least 4 orders of magnitude, meaning that samples do not have to contain equal starting amounts of DNA. This technique should therefore be suitable for analyzing formalin-fixed, paraffin-embedded tissues. Fourth, and above all, real-time PCR makes DNA quantification much more precise and reproducible, since it is based on  $C_t$  values rather than end-point measurement of the amount of accumulated PCR product. Indeed, the ABI Prism 7700 Sequence Detection System enables  $C_t$  to be calculated when PCR amplification is still in the exponential phase and when none of the reaction components is rate-limiting. The within-run CV of the  $C_t$  value for calibrator human DNA (5 replicates) was always below 5%, and the between-assay precision in 5 different runs was always below 10% (data not shown). In addition, the use of a standard curve is not absolutely necessary, since the copy number can be determined simply by comparing the  $C_t$  ratio of the target gene with that of reference genes. The results obtained by the 2 methods (with and without a standard curve) are similar in our experiments (data not shown). Moreover, unlike competitive quantitative PCR, real-time PCR does not require an internal control (the design and storage of internal controls and the validation of their amplification efficiency is laborious).

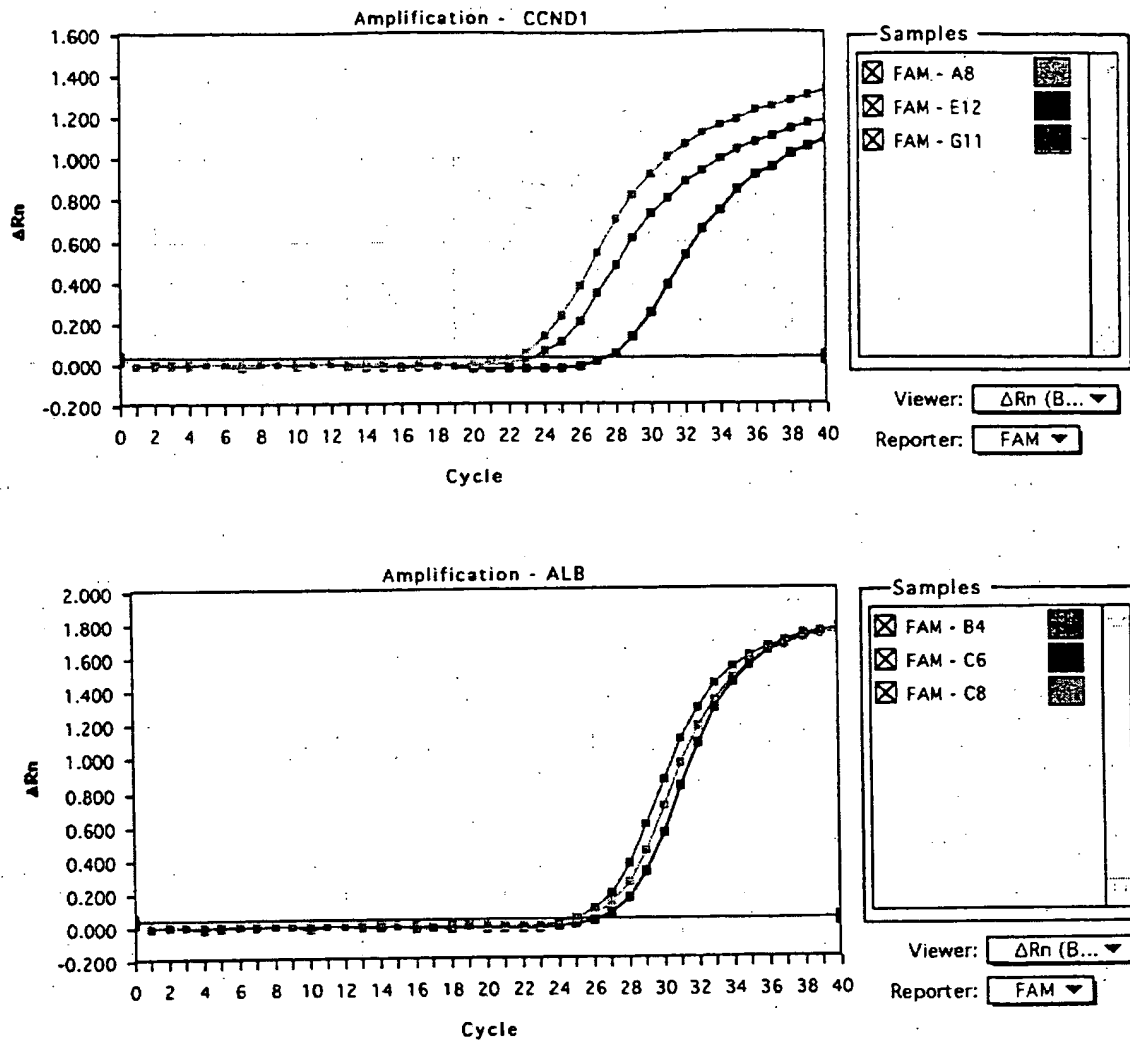
The only potential disadvantage of real-time PCR, like all other PCR-based methods and solid-matrix blotting techniques (Southern blots and dot blots) is that it cannot avoid dilution artifacts inherent in the extraction of DNA from tumor cells contained in heterogeneous tissue specimens. Only FISH and immunohistochemistry can measure alterations on a cell-by-cell basis (Pauletti *et al.*, 1996; Slamon *et al.*, 1989). However, FISH requires expensive equipment and trained personnel and is also time-consuming. Moreover, FISH does not assess gene expression and therefore cannot detect cases in which the gene product is over-expressed in the absence of gene amplification, which will be possible in the future by real-time quantitative RT-PCR. Immunohistochemistry is subject to considerable variations in the hands of different teams, owing to alterations of target proteins during the procedure, the different primary antibodies and fixation methods used and the criteria used to define positive staining.

The results of this study are in agreement with those reported in the literature. (i) Chromosome regions 4q11-q13 and 21q21.2 (which bear *alb* and *app*, respectively) showed no genetic alterations in the breast-cancer samples studied here, in keeping with the results of CGH (Kallioniemi *et al.*, 1994). (ii) We found that amplifications of these 3 oncogenes were independent events, as reported by other teams (Berns *et al.*, 1992; Borg *et al.*, 1992). (iii) The frequency and degree of *myc* amplification in our breast tumor DNA series were lower than those of *ccnd1* and *erbB2* amplification, confirming the findings of Borg *et al.* (1992) and Courjal *et al.* (1997). (iv) The maxima of *ccnd1* and *erbB2* over-representation were 18-fold and 15-fold, also in keeping with earlier results (about

TABLE I - DISTRIBUTION OF AMPLIFICATION LEVEL (N) FOR *myc*, *ccnd1* AND *erbB2* GENES IN 108 HUMAN BREAST TUMORS

Gene	Amplification level (N)			
	<0.5	0.5-1.9	2-4.9	$\geq 5$
<i>myc</i>	0	97 (89.8%)	11 (10.2%)	0
<i>ccnd1</i>	0	83 (76.9%)	17 (15.7%)	8 (7.4%)
<i>erbB2</i>	5 (4.6%)	87 (80.6%)	8 (7.4%)	8 (7.4%)





Tumor	CCND1		ALB	
	$C_t$	Copy number	$C_t$	Copy number
■ T118	27.3	4605	26.5	4365
■ T133	23.2	61659	25.2	10092
■ T145	22.1	125892	25.6	7762

FIGURE 2 – *ccnd1* and *alb* gene dosage by real-time PCR in 3 breast tumor samples: T118 (E12, C6, black squares), T133 (G11, B4, red squares) and T145 (A8, C8, blue squares). Given the  $C_t$  of each sample, the initial copy number is inferred from the standard curve obtained during the same experiment. Triplicate plots were performed for each tumor sample, but the data for only one are shown here. The results are shown in Table II.

30-fold maximum) (Berns *et al.*, 1992; Borg *et al.*, 1992; Courjal *et al.*, 1997). (v) The *erbB2* copy numbers obtained with real-time PCR were in good agreement with data obtained with other quantitative PCR-based assays in terms of the frequency and degree of amplification (An *et al.*, 1995; Deng *et al.*, 1996; Valeron

*et al.*, 1996). Our results also correlate well with those recently published by Gelmini *et al.* (1997), who used the TaqMan system to measure *erbB2* amplification in a small series of breast tumors ( $n = 25$ ), but with an instrument (LS-50B luminescence spectrometer, Perkin-Elmer Applied Biosystems) which only allows end-

TABLE II - EXAMPLES OF *ccnd1* GENE DOSAGE RESULTS FROM 3 BREAST TUMORS<sup>1</sup>

Tumor	<i>ccnd1</i>			<i>alb</i>			<i>Nccnd1/alb</i>
	Copy number	Mean	SD	Copy number	Mean	SD	
T118	4525	4603	77	4223	4325	89	1.06
	4605			4365			
	4678			4387			
T133	59821	61100	1111	9787	10137	375	6.03
	61659			10092			
	61821			10533			
T145	128563	125392	3448	7321	7672	316	16.34
	125892			7762			
	121722			7933			

<sup>1</sup>For each sample, 3 replicate experiments were performed and the mean and the standard deviation (SD) was determined. The level of *ccnd1* gene amplification (*Nccnd1/alb*) is determined by dividing the average *ccnd1* copy number value by the average *alb* copy number value.

point measurement of fluorescence intensity. Here we report *myc* and *ccnd1* gene dosage in breast cancer by means of quantitative PCR. (vi) We found a high degree of concordance between real-time quantitative PCR and Southern blot analysis in terms of gene amplification, especially for samples with high copy numbers ( $\geq 5$ -fold). The slightly higher frequency of gene amplification (especially *ccnd1* and *erbB2*) observed by means of real-time quantitative PCR as compared with Southern-blot analysis may be explained by the higher sensitivity of the former method. However, we cannot rule out the possibility that some tumors with a few extra

gene copies observed in real-time PCR had additional copies of an arm or a whole chromosome (trisomy, tetrasomy or polysomy) rather than true gene amplification. These 2 types of genetic alteration (polysomy and gene amplification) could be easily distinguished in the future by using an additional probe located on the same chromosome arm, but some distance from the target gene. It is noteworthy that high gene copy numbers have the greatest prognostic significance in breast carcinoma (Borg *et al.*, 1992; Slamon *et al.*, 1987).

Finally, this technique can be applied to the detection of gene deletion as well as gene amplification. Indeed, we found a decreased copy number of *erbB2* (but not of the other 2 proto-oncogenes) in several tumors; *erbB2* is located in a chromosome region (17q21) reported to contain both deletions and amplifications in breast cancer (Bièche and Lidereau, 1995).

In conclusion, gene amplification in various cancers can be used as a marker of pre-neoplasia, also for early diagnosis of cancer, staging, prognostication and choice of treatment. Southern blotting is not sufficiently sensitive, and FISH is lengthy and complex. Real-time quantitative PCR overcomes both these limitations, and is a sensitive and accurate method of analyzing large numbers of samples in a short time. It should find a place in routine clinical gene dosage.

#### ACKNOWLEDGEMENTS

RL is a research director at the Institut National de la Santé et de la Recherche Médicale (INSERM). We thank the staff of the Centre René Huguenin for assistance in specimen collection and patient care.

#### REFERENCES

- AN, H.X., NIEDERACHER, D., BECKMANN, M.W., GÖHRING, U.J., SCHARL, A., PICARD, F., VAN ROEYEN, C., SCHNÜRCH, H.G. and BENDER, H.G., *erbB2* gene amplification detected by fluorescent differential polymerase chain reaction in paraffin-embedded breast carcinoma tissues. *Int. J. Cancer (Pred. Oncol.)*, 64, 291-297 (1995).
- BERNS, E.M.J.J., KLIJN, J.G.M., VAN PUTTEN, W.L.J., VAN STAVEREN, I.L., PORTINGEN, H. and FOEKENS, J.A., *c-myc* amplification is a better prognostic factor than *HER2/neu* amplification in primary breast cancer. *Cancer Res.*, 52, 1107-1113 (1992).
- BIÈCHE, I. and LIDEREAU, R., Genetic alterations in breast cancer. *Genes Chrom. Cancer*, 14, 227-251 (1995).
- BORG, A., BALDETORP, B., FERNO, M., OLSSON, H. and SIGURDSSON, H., *c-myc* amplification is an independent prognostic factor in post-menopausal breast cancer. *Int. J. Cancer*, 51, 687-691 (1992).
- CELI, F.S., COHEN, M.M., ANTONARAKIS, S.E., WERTHEIMER, E., ROTH, J. and SHULDINER, A.R., Determination of gene dosage by a quantitative adaptation of the polymerase chain reaction (qd-PCR): rapid detection of deletions and duplications of gene sequences. *Genomics*, 21, 304-310 (1994).
- COURJAL, F., CUNY, M., SIMONY-LAFONTAINE, J., LOUASSON, G., SPEISER, P., ZEILLINGER, R., RODRIGUEZ, C. and THEILLET, C., Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.*, 57, 4360-4367 (1997).
- DENG, G., YU, M., CHEN, L.C., MOORE, D., KURISU, W., KALLIONIEMI, A., WALDMAN, F.M., COLLINS, C. and SMITH, H.S., Amplifications of oncogene *erbB-2* and chromosome 20q in breast cancer determined by differentially competitive polymerase chain reaction. *Breast Cancer Res. Treat.*, 40, 271-281 (1996).
- GELMINI, S., ORIANDO, C., SESTINI, R., VONA, G., PINZANI, P., RUOCCO, L. and PAZZAGLI, M., Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure *c-erbB-2* oncogene amplification. *Clin. Chem.*, 43, 752-758 (1997).
- GIBSON, U.E.M., HEID, C.A. and WILLIAMS, P.M., A novel method for real-time quantitative RT-PCR. *Genome Res.*, 6, 995-1001 (1996).
- HEID, C.A., STEVENS, J., LIVAK, K.J. and WILLIAMS, P.M., Real-time quantitative PCR. *Genome Res.*, 6, 986-994 (1996).
- HOLLAND, P.M., ABRAMSON, R.D., WATSON, R. and GELFAND, D.H., Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. nat. Acad. Sci. (Wash.)*, 88, 7276-7280 (1991).
- KALLIONIEMI, A., KALLIONIEMI, O.P., PIPER, J., TANNER, M., STOKKES, T., CHEN, L., SMITH, H.S., PINKEL, D., GRAY, J.W. and WALDMAN, F.M., Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. nat. Acad. Sci. (Wash.)*, 91, 2156-2160 (1994).
- LEE, L.G., CONNELL, C.R. and BIOCH, W., Allelic discrimination by nick-translation PCR with fluorogenic probe. *Nucleic Acids Res.*, 21, 3761-3766 (1993).
- LONGO, N., BERNINGER, N.S. and HARTLEY, J.L., Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene*, 93, 125-128 (1990).
- MUSS, H.B., THOR, A.D., BERRY, D.A., KUTE, T., LIU, E.T., KOERNER, F., CIRRIKIONE, C.T., BUDMAN, D.R., WOOD, W.C., BARCOS, M. and HENDERSON, I.C., *c-erbB-2* expression and response to adjuvant therapy in women with node-positive early breast cancer. *New Engl. J. Med.*, 330, 1260-1266 (1994).
- PAULETTI, G., GODOLPHIN, W., PRESS, M.F. and SALMON, D.J., Detection and quantification of *HER-2/neu* gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene*, 13, 63-72 (1996).
- PIATAK, M., LUK, K.C., WILLIAMS, B. and LIFSON, J.D., Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Biotechniques*, 14, 70-80 (1993).
- SCHUURING, E., VERHOEVEN, E., VAN TINTEREN, H., PETERSE, J.L., NUNNIK, B., THUNNISSEN, F.B.J.M., DEVILLE, P., CORNELISSE, C.J., VAN DE VIVER, M.J., MOOI, W.J. and MICHALIDES, R.J.A.M., Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. *Cancer Res.*, 52, 5229-5234 (1992).
- SLAMON, D.J., CLARK, G.M., WONG, S.G., LEVIN, W.S., ULLRICH, A. and MCGUIRE, W.L., Human breast cancer: correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science*, 235, 177-182 (1987).
- SLAMON, D.J., GODOLPHIN, W., JONES, L.A., HOLT, J.A., WONG, S.G., KEITH, D.E., LEVIN, W.J., STUART, S.G., UDOWE, J., ULLRICH, A. and PRESS, M.F., Studies of the *HER-2/neu* proto-oncogene in human breast and ovarian cancer. *Science*, 244, 707-712 (1989).
- VALERON, P.F., CHIRINO, R., FERNANDEZ, L., TORRES, S., NAVARRO, D., AGUIAR, J., CABRERA, J.J., DIAZ-CHICO, B.N. and DIAZ-CHICO, J.C., Validation of a differential PCR and an ELISA procedure in studying *HER-2/neu* status in breast cancer. *Int. J. Cancer*, 65, 129-133 (1996).

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Ashkenazi et al.  
App. No. : 09/903,925  
Filed : July 11, 2001  
For : SECRETED AND  
TRANSMEMBRANE  
POLYPEPTIDES AND NUCLEIC  
ACIDS ENCODING THE SAME  
Examiner : Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

(Date)

Commissioner of Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

**DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132**

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.
2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan<sup>®</sup> PCR or the fluorescence *in situ* hybridization (FISH) assays - is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi  
Avi Ashkenazi, Ph.D.

Date: 9/15/03

## **CURRICULUM VITAE**

**Avi Ashkenazi**

July 2003

### **Personal:**

Date of birth: 29 November, 1956  
Address: 1456 Tarrytown Street, San Mateo, CA 94402  
Phone: (650) 578-9199 (home); (650) 225-1853 (office)  
Fax: (650) 225-6443 (office)  
Email: aa@gene.com

### **Education:**

1983: B.S. in Biochemistry, with honors, Hebrew University, Israel  
1986: Ph.D. in Biochemistry, Hebrew University, Israel

### **Employment:**

1983-1986: Teaching assistant, undergraduate level course in Biochemistry  
1985-1986: Teaching assistant, graduate level course on Signal Transduction  
1986 - 1988: Postdoctoral fellow, Hormone Research Dept., UCSF, and  
Developmental Biology Dept., Genentech, Inc., with J. Ramachandran  
1988 - 1989: Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc.,  
with D. Capon  
1989 - 1993: Scientist, Molecular Biology Dept., Genentech, Inc.  
1994 -1996: Senior Scientist, Molecular Oncology Dept., Genentech, Inc.  
1996-1997: Senior Scientist and Interim director, Molecular Oncology Dept.,  
Genentech, Inc.  
1997-1990: Senior Scientist and preclinical project team leader, Genentech, Inc.  
1999 -2002: Staff Scientist in Molecular Oncology, Genentech, Inc.  
2002-present: Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

### **Awards:**

1988: First prize, The Boehringer Ingelheim Award

## Editorial:

Editorial Board Member: Current Biology

Associate Editor, Clinical Cancer Research.

Associate Editor, Cancer Biology and Therapy.

## Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* **34**, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* **118**, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* **50**, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internatl.* **14**, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* **210**, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on in vitro and in vivo modulation of lactogenic and somatotrophic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* **121**, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* **236**, 600-605 (1987).
8. Peralta, E., Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* **6**, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* **238**, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* **37**, 119-129 (1988).
11. Peralta, E., Ashkenazi, A., Winslow, J., Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* **334**, 434-437 (1988).
12. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* **56**, 487-493 (1989).
13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* **340**, 146-150 (1989).
14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* **245**, 743-745 (1989).
15. Ashkenazi, A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodeficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* **87**, 7150-7154 (1990).
16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* **29**, 9885-9891 (1990).
17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* **88**, 7056-7060 (1991).
18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari, I., Pennica, D., Goeddel, D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* **88**, 10535-10539 (1991).
19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* **66**, 235-243 (1992).



20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel, R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine<sub>1B</sub> receptor. *J. Biol. Chem.* **267**, 5735-5738 (1992).
21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* **267**, 5747-5750 (1992).
22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* **267**, 15916-15922 (1992).
23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT<sub>1B</sub> receptors. *Nature* **360**, 161-163 (1992).
24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon  $\gamma$  by an interferon  $\gamma$  receptor immunoadhesin. *Immunology* **79**, 594-599 (1993).
25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, A., Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* **32**, 3131-3138. (1993).
26. Barfod, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* **268**, 26059-26062 (1993).
27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3<sup>+</sup> effectors to kill HIV-1-infected cells. *J. Immunol.* **153**, 4268-4280 (1994).
28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by iterferon  $\gamma$  and correction by interferon  $\gamma$  receptor immunoadhesin. *Blood* **83**, 911-915 (1994).
29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., and Ashkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* **152**, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkylation. *Bioconj. Chem.* **5**, 133-140 (1994).
31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* **170**, 1323-1326 (1994).
32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* **31**, 1335-1344 (1994).
33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* **31**, 1345-1351 (1994).
34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT<sub>2</sub> receptor specifies effector coupling. *J. Neurochem.* **64**, 1440-1447 (1995).
35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN- $\gamma$  receptor  $\beta$  chain expression in T helper cell subsets. *Science* **270**, 1215-1218 (1995).
36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* **95**, 1947-1952 (1995).
37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon  $\gamma$  signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA*. **92**, 5401-5405 (1995).
38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal *E. coli* bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* **156**, 2221-2230 (1996).
39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* **271**, 12687-12690 (1996).

40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 1669-1676 (1996).
41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF- $\kappa$ B and AP-1 transcription factors. *J. Biol. Chem.* 272, 14029-14032 (1997).
42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* 277, 818-821 (1997).
43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* 7, 1003-1006 (1997).
44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* 427, 124-128 (1998).
46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* 160, 4098-4103 (1998).
47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* 396, 699-703 (1998).
48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* 162, 5616-5623 (1999).
49. Gurney, A. Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* 9, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, s., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* 104, 155-162 (1999).
51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmissiion in the human IFN- $\alpha$  receptor 2. *J. Immunol.* 163, 766-773 (1999).
52. Johnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* 11, 664-672 (1999).
53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* 265, 479-483 (1999).
54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* 4, 563-571 (1999).
55. Hymowitz, S.G., O'Connel, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* 39, 633-640 (2000).
56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* 60, 2611-2615 (2000).
57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* 12, 611-620 (2000).
58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., \*Ashkenazi, A., and \*Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* 1, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* 10, 785-788 (2000).
60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* 29, 506-512 (2000).
61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D. Mounho, B., Hillan, K., Totpal, K. DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* 7, 383-385 (2001).
62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Marsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* 166, 4891-4898 (2001).
63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* 7, 1362-1369 (2001).
64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., \*Ashkenazi, A., and \*Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* 2, 632-637 (2001).
65. Kischkel, F.C., Lawrence, D. A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* 276, 46639-46646 (2001).
66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A. Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotic Bcl-2 homolog Bax. *Nature Med.* 8, 274-281 (2002).
67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault, G., Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* 170, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

**Review articles:**

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology*. **LIII**, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliege, A., Ammann, A., Kahn, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* **10**, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels*. (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. I, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. 3, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* **1**, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzymology* **8**, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* **14**, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* **281**, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* **11**, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook ([www.apnet.com/cytokinereference/](http://www.apnet.com/cytokinereference/)).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* 2, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. *Cytokine and Growth Factor Reviews* 14, 337-348 (2003).

**Book:**

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

**Talks:**

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Half-life Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoadhesins for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June 1992.
4. Immunoadhesins: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficacy of TNF receptor immunoadhesin vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon- $\gamma$  signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Francisco, CA, July 1995.
8. Immunoadhesins: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: a novel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philadelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.



25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. . Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philladelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14<sup>th</sup> Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Weizmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA. October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Switzerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

**Issued Patents:**

1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6,046,048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).

## Analysis of Genomic and Proteomic Data Using Advanced Literature Mining

Yanhui Hu, Lisa M. Hines, Haifeng Weng, Dongmei Zuo, Miguel Rivera,  
Andrea Richardson, and Joshua LaBaer\*

*Institute of Proteomics, Harvard Medical School-BCMP, 240 Longwood Avenue, Boston, Massachusetts 02115*

Received March 13, 2003

High-throughput technologies, such as proteomic screening and DNA micro-arrays, produce vast amounts of data requiring comprehensive analytical methods to decipher the biologically relevant results. One approach would be to manually search the biomedical literature; however, this would be an arduous task. We developed an automated literature-mining tool, termed MedGene, which comprehensively summarizes and estimates the relative strengths of all human gene-disease relationships in Medline. Using MedGene, we analyzed a novel micro-array expression dataset comparing breast cancer and normal breast tissue in the context of existing knowledge. We found no correlation between the strength of the literature association and the magnitude of the difference in expression level when considering changes as high as 5-fold; however, a significant correlation was observed ( $r = 0.41$ ;  $p = 0.05$ ) among genes showing an expression difference of 10-fold or more. Interestingly, this only held true for estrogen receptor (ER) positive tumors, not ER negative. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative tumors worthy of further examination.

**Keywords:** bioinformatics • micro-array • text mining • gene-disease association • breast cancer

### Introduction

At its current pace, the accumulation of biomedical literature outpaces the ability of most researchers and clinicians to stay abreast of their own immediate fields, let alone cover a broader range of topics. For example, to follow a single disease, e.g., breast cancer, a researcher would have had to scan 130 different journals and read 27 papers per day in 1999.<sup>1</sup> This problem is accentuated with high-throughput technologies such as DNA micro-arrays and proteomics, which require the analysis of large datasets involving thousands of genes, many of which are unfamiliar to a particular researcher. In any microarray experiment, thousands of genes may demonstrate statistically significant expression changes, but only a fraction of these may be relevant to the study. The ability to interpret these datasets would be enhanced if they could be compared to a comprehensive summary of what is known about all genes. Thus, there is a need to summarize existing knowledge in a format that allows for the rapid analysis of associations between genes and diseases or other specific biological concepts.

One solution to this problem is to compile structured digital resources, such as the Breast Cancer Gene Database<sup>1</sup> and the Tumor Gene Database.<sup>2</sup> However, as these resources are hand-curated, the labor-intensive review process becomes a rate-limiting step in the growth of the database. As a result, these

databases have a limited scale and the genes are not selected in a systematic fashion.

An alternative approach is automated text mining: a method which involves automated information extraction by searching documents for text strings and analyzing their frequency and context. This approach has been used successfully in several instances for biological applications. In most cases, it has been applied to extract information about the relationships or interactions that proteins or genes have with one another, in the literature or by functional annotation.<sup>3-7</sup> Thus far, few publications have applied text-mining to examine the global relationships between genes and diseases. Perez-Iratxeta et al. automatically examined the GO (Gene Ontology) annotation of genes and their predicted chromosomal locations in order to identify genes linked to inherited disorders.<sup>8</sup>

To obtain a more global understanding of disease development, it would be valuable to incorporate information regarding all possible gene-disease relationships, including biochemical, physiological, pharmacological, epidemiological, as well as genetic. This information would enable comprehensive comparisons between large experimental datasets and existing knowledge in the literature. This would accomplish two things. First, it would serve to validate experiments by demonstrating that known responses occur as predicted. Second, it would rapidly highlight which genes are corroborated by the literature and which genes are novel in a given context. We have utilized a computational approach to literature mining to produce a

\* To whom correspondence should be addressed: jlabae@hms.harvard.edu.

comprehensive set of gene-disease relationships. In addition, we have developed a novel approach to assess the strength of each association based on the frequency of citation and co-citation. We applied this tool to help interpret the data from a large micro-array gene expression experiment comparing normal and cancerous breast tissue.

## Methods

**MedGene Database.** MedGene is a relational database, storing disease and gene information from NCBI, text mining results, statistical scores, and hyperlinks to the primary literature. MedGene has a web-based user interface for users to query the database (<http://hipseq.med.harvard.edu/MedGene/>).

**Text Mining Algorithms.** MeSH files were downloaded from the MeSH web site at NLM (National Library of Medicine) (<http://www.nlm.nih.gov/mesh/meshhome.html>) and human disease categories were selected. LocusLink files were downloaded from the LocusLink web site at NCBI (<http://www.ncbi.nih.gov/LocusLink/>). Official/preferred gene symbol, official/preferred gene name, and gene alternative symbols and names, all relevant annotations and URLs for each LocusLink record, were collected. Gene search terms were used for literature searching and included all qualified gene names, gene symbols, and gene family terms. Primary gene keys, predominantly qualified gene family terms and gene official/preferred symbols, were used to index Medline records. If the official/preferred gene symbols did not meet the standards to be an index, then qualified gene official/preferred names were used. A local copy of Medline records (up to July, 2002) was pre-selected.

A JAVA module examined the MeSH terms and then indexed each Medline record with the appropriate disease terms. A separate JAVA module was used to examine the titles and abstracts for gene search terms and then to index the gene-related Medline records with the relevant primary gene key(s).

**Statistical Methods.** For every gene and disease pair, we counted records that were indexed for both gene and disease (double positive hits), for disease only (disease single hits), for gene only (gene single hits), and for neither gene nor disease (double negative hits) to generate a  $2 \times 2$  contingency table. On the basis of the contingency table framework, we applied different statistical methods to estimate the strength of gene-disease relationships and evaluated the results. These methods included chi-square analysis, Fisher's exact probabilities, relative risk of gene, and relative risk of disease<sup>10</sup> (<http://hipseq.med.harvard.edu/MedGene/>). In addition, we computed the "product of frequency", which is the product of the proportion of disease/gene double hits to disease single hits and the proportion of disease/gene double hits to gene single hits. To obtain a normal distribution, we transformed all the statistical scores using the natural logarithm. We selected the log of the product of frequency (LPF) to validate MedGene and to use for the analysis with the micro-array data. Spearman rank-correlation coefficients were used to assess the linear relationship between LPF and micro-array fold change in expression level.

**Global Analysis.** Diseases with at least 50 related genes were selected for clustering analysis, and the LPF scores were normalized with total score for each disease. Hierarchical clustering was done with the "Cluster" software and the clustering result was visualized using "TreeView" (<http://rana.lbl.gov/EisenSoftware.htm>).

**Breast Tissue Micro-Arrays.** Eighty-nine breast cancer samples (79% ER-positive) and 7 normal breast tissue samples were selected from the Harvard Breast SPORE frozen tissue repository and were representative of the spectrum of histological types, grades, and hormone receptor immunophenotypes of breast cancer. Biotinylated cRNA, generated from the total RNA extracted from the bulk tumor, was hybridized to Affymetrix U95A oligo-nucleotide micro-arrays. These micro-arrays consist of 12 400 probes, which represent approximately 9000 genes. Raw expression values were obtained using GENE-CHIP software from Affymetrix, and then further analyzed using the DNA-Chip Analyzer (dChip) custom software.

## Results

**Automated Indexing of Medline Records by Disease and Gene.** To study the gene-disease associations in the literature, we first compiled complete lists for human diseases and human genes. To index all Medline records that were relevant to human diseases, the Medical Subject Heading (MeSH) index of Medline records was utilized. MeSH is a controlled medical vocabulary from the National Library of Medicine and consists of a set of terms or subject headings that are arranged in both an alphabetic and an hierarchical structure. Medline records are reviewed manually and MeSH terms are added to each with software assistance.<sup>9,10</sup> Twenty-three human disease category headings along with all of their child terms (see the Supporting Information, Supplemental Table 1, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table1.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table1.html)) were selected from the 2002 MeSH index creating a list of 4033 human diseases.

No index comparable to the MeSH index exists for genes, and thus, it was necessary to apply a string search algorithm for gene names or symbols found in Medline text. A complete list of genes, gene names, gene symbols, and frequently used synonyms were collected from the LocusLink database at NCBI,<sup>11,12</sup> which contains 53 259 independent records keyed by an official gene symbol or name (June 18<sup>th</sup>, 2002). For the purposes of this study, no distinction was made between genes and their gene products. Authors often use the same name for both, differentiating the two only by the use of italics, if at all. For the intended use of this study, this lack of distinction is unlikely to have a large effect and may in fact be beneficial.

Initial attempts to search the literature using these lists revealed several sources of false positives and false negatives (Table 1). False positives primarily arose when the searched term had other meanings, whereas false negatives arose from syntax discrepancies necessitating the development of filters to reduce these errors. The syntax issues were readily handled by including alternate syntax forms in the search terms. The false positive cases, caused by duplicative and unrelated meanings for the terms, were more difficult to manage. Where possible, case sensitive string mapping reduced inappropriate citations. In many cases, however, this was not sufficient and the terms had to be eliminated entirely, thereby reducing the false positive rate but unavoidably under-representing some genes.

For the purposes of data tracking, a primary gene key was selected to represent all synonyms that correspond to each gene. Medline records were indexed with a primary gene key when any synonym for that key was found in the title or abstract. Case-insensitive string mapping was used for all searches except as noted above. No additional weight was

Table 1. Systematic Sources of False Positives and False Negatives in Unfiltered Data\*

source of error	error type	example	filter solution
gene symbol/name is not unique	false positive	MAC—myelin associated glycoprotein MAC—malignancy-associated protein	eliminate this term
gene symbol is unrelated abbreviation	false positive	PA—pallid homologue (mouse), pallidin (also abbrev. for Pennsylvania)	eliminate this term
gene symbol/name has language meaning	false positive	WAS—Wiskott—Aldrich Syndrome (also the word "was")	case-sensitive string search
nonstandard syntax	false negative	BAG-1 instead of BAG1	add dash term
unofficial gene name/symbol	false negative	P53 instead of TP53	add all gene nicknames
nonspecified gene name	false negative	estrogen receptor instead of Estrogen receptor 1	add family stem term

\* In preliminary studies, Medline was searched for co-occurrence of genes and diseases and the resulting output was evaluated to identify error sources that were amenable to global filters. Each error source is categorized by the type of error it causes: false positives are suggested relationships that are not real and false negatives are real relationships that are underrepresented. The filter solutions used are indicated. Note that in some cases, the filter solution itself introduces error. In general, error rates maximized sensitivity, even at the expense of specificity if needed.

added for multiple occurrences of a term or the co-occurrence of multiple synonyms for the same gene key.

Medline records were searched with all qualified gene identifiers, such as the official/preferred gene symbol, the official/preferred gene name, all gene nicknames and all syntax variants. In situations where there are several members of a gene family or splice variants, some authors prefer to use a shortened gene family name, e.g., estrogen receptor, instead of estrogen receptor 1 (*ESR1*), creating a source of false negatives. For this reason, gene family stem terms were created for all genes that have an alpha or numerical suffix (e.g., *IL2RA*, *TGFB*, *ESR1*, etc.) and then used to search the literature. The family stem terms were handled separately from the specific gene names so that it would be clear when linkages were made to the gene family versus a specific member in that family.

To improve performance and accuracy, some pre-selection was applied to the records that were scanned. First, review articles were eliminated to avoid redundant treatment of citations. Second, non-English journals were removed because the natural language filters were only relevant to English publications. Finally, journals unlikely to contain primary data about gene-disease relationships were also removed (e.g., *Int. J. Health Educ.*, *Bedside Nurse*, and *J. Health Econ.*). Together, these filters reduced the 12 198 221 Medline publications (July 2002) by 37%.

**Ranking the Relative Strengths of Gene-Disease Associations.** In total, there were 618 708 gene-disease co-citations. In which 16% (8297) of all studied genes had been associated to a disease and 96% (3875) of all diseases had been associated to at least one gene. To rank the relative strengths of gene disease relationships, we tested several different statistical methods and examined the results. With the exception of the relative risk estimates, the methods provided similar results with respect to the rank order of the gene-disease association strengths. However, after comparing the results to other databases and after consulting disease experts, the log of the product of frequency (LPF) was selected for further analysis because it gave the best results overall.

**Validation of MedGene.** In developing this tool, it was important to minimize the number of missed genes (false negatives) and misclassified genes (false positives). However, in situations when these goals were in conflict, inclusiveness was prioritized. To determine the false negative rate in MedGene, breast cancer was used as a test case because it was associated with more genes than any other human disease and because

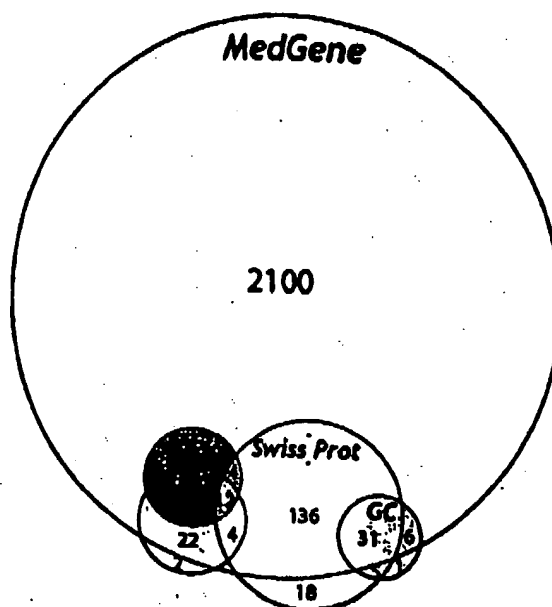


Figure 1. Estimation of the false negative rate by comparison with hand-curated databases. The breast cancer-related genes identified by MedGene were compared with those listed in several other databases including the Tumor Gene Database (TGD),<sup>2</sup> the Breast Cancer Gene Database (BCG),<sup>1</sup> GeneCards (GC)<sup>17</sup> and Swissprot.<sup>18</sup> Genes were considered false negatives if they were represented in at least one of these other databases and not in MedGene and their link to breast cancer was supported by at least one literature reference. All literature references were verified by manual review to confirm their validity. The number of genes in each database or shared by more than one database is indicated. The false negative rate was calculated by genes missed at MedGene (26)/total number of nonoverlapping genes in other databases (285).

there were several public databases that link genes to breast cancer. We compared the list of breast cancer-related genes from MedGene to these databases. Illustrated in Figure 1. Among the 285 distinct breast cancer-related genes that were supported by at least one literature citation in these hand-curated databases, 26 were absent from MedGene, suggesting a false negative rate of approximately 9%. To determine why these were missed, all literature references for these genes (80

papers) were reviewed manually (see the Supporting Information, Supplemental Table 2, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table\\_2.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table_2.html)). Among these papers, most false negatives were caused by nonstandard gene terms or gene terms eliminated by our specificity filters. Few genes were missed because they were only mentioned in review papers (0.4%) or they appeared only in the body of the manuscript but not the abstract or title (1.1%). Of note, MedGene identified approximately 2000 additional breast cancer-related genes not listed in any other database.

To assess the false positive error rate, two complementary approaches were used: a detailed analysis of one disease and a global examination of 1000 diseases. The detailed approach examined the false positive error rate and its sources, whereas the global approach tested whether the overall results made biomedical sense.

Using the LPP, 1467 genes related to prostate cancer were assembled in rank order. We then retrieved approximately 300 Medline records each for the highest ranked 100 and the lowest ranked 200 genes and manually reviewed the titles and abstracts to determine the verity of the association. Nearly 80% of the highest ranked 100 genes fell into one of the five categories that reflect meaningful gene-disease relationships (see the Supporting Information, Supplemental Table 3, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table\\_3.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table_3.html)). Among the lowest ranked 200 genes, approximately 70% reflected true relationships. Of the 600 records reviewed, there were only two in which the association between the gene and the disease was described as negative. Both were genes with very low scores. In both cases, the authors did not argue the absence of any relationship, but rather that a particular feature of the gene or protein was not shown to be related to human prostate cancer.<sup>12,14</sup>

The coincidence of some gene symbols with medical abbreviations, chemical abbreviations and biological abbreviations resulted in most of the false positives (see the Supporting Information, Supplemental Table 4, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table\\_4.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table_4.html)), emphasizing the importance of the filters that were added in the search algorithm (Table 1). Without the filters, the false positive rate more than doubled, and the false negative rate rose dramatically (data not shown). For example, among the papers about breast cancer, there were only 12 Medline records that referred to *ESR1* and 10 to *ESR2*, whereas almost 2000 papers mentioned estrogen receptor without specifying *ESR1* or *ESR2*; this latter group was detected by the family term filter.

To further validate these results, a global analysis of the gene-disease relationships described by MedGene was performed. For this experiment, it was reasoned that the more closely related the diseases are to one another, the more they will be related to the same gene sets. Thus, if the relationships defined by MedGene accurately reflected the literature, then an unsupervised hierarchical clustering of the gene data should group diseases in a manner consistent with common medical thinking. Conversely, if the clustered diseases do not make sense biologically or medically, it may reflect excessive false positives, false negatives, or inappropriate scoring of the data.

To execute this experiment, the gene sets and the corresponding LPP values for 1000 randomly selected diseases (each with at least 50 gene relationships) were used as a dataset for clustering the diseases. A review of the results showed that the resulting disease clusters were indeed logical based upon common medical knowledge (see the Supporting Information,

Supplemental Figure 1, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Figure\\_1.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_1.html)). For example, in one such cluster shown in Figure 2, diabetes and its complications grouped together and were also closely linked to diseases associated with starvation states.

The number of genes associated with a given disease can be estimated by adjusting the MedGene number up by the false negative rate (~9%) and down by the false positive rate (~26% on average). Using this, the average disease has  $103.7 \pm 45.3$  (mean  $\pm$  s.d.) genes associated with it, although the range is quite broad with 2359 genes related to breast cancer, 2122 genes related to lung cancer and no genes related to a number of diseases.

**Applying MedGene to the Analysis of Large Datasets.** Access to a comprehensive summary of the genes linked to human diseases provided an opportunity to analyze data obtained from a high-throughput experiment. We compared the MedGene breast cancer gene list to a gene expression data set generated from a micro-array analysis comparing breast cancer and normal breast tissue samples. Micro-array analysis identified 2286 genes that had greater than a 1-fold difference in mean expression level between breast cancer samples and normal breast samples. Using MedGene, we sorted the 2286 genes into four classes: 555 genes directly linked to breast cancer in the literature by gene term search (first-degree association by gene name); 328 genes directly linked by family term search (first-degree association by family term); 1021 genes linked to breast cancer only through other breast cancer genes (second-degree association); and 505 genes not previously associated with breast cancer. (See the Supporting Information, Supplemental Figure 2, or visit [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Figure\\_2.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_2.html).) Among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease. Among the remaining 38 genes, 9 had been related to other cancers, specifically esophageal, colon, uterine, skin, and cervix.

To determine whether the genes highlighted by the micro-array analysis were more likely to have been previously linked to breast cancer in the literature, we created a two-dimensional plot of the fold change of expression level between breast cancer and normal tissue versus the literature score (LPP) (Figure 3A). There was a broad spread of expression changes among the genes directly linked to breast cancer ranging from less than 1-fold change (68%) to over 40-fold (0.3%). Notably, the majority of genes with greater than 10-fold expression changes were linked to breast cancer by first-degree association.

Among all 754 genes directly linked to breast cancer in the literature, there was no correlation between LPP and micro-array fold change ( $r = 0.018$ ,  $p$ -value = 0.62). However, when we stratified the analysis based on the magnitude of the fold change, we observed an increasing trend in correlation (Figure 3B) suggesting that genes with a more substantial change in expression level were more likely to have a stronger association in the literature. For genes that had 10-fold change or more in expression level, the correlation increased to 0.41 ( $p$ -value = 0.05).

When we evaluated the micro-array data separately for ER positive and ER negative tumors, the trend in correlation between fold change and literature score was highly dependent on estrogen receptor status. Interestingly, there was a similar trend in correlation for ER positive tumors, but no trend in correlation for ER negative tumors.

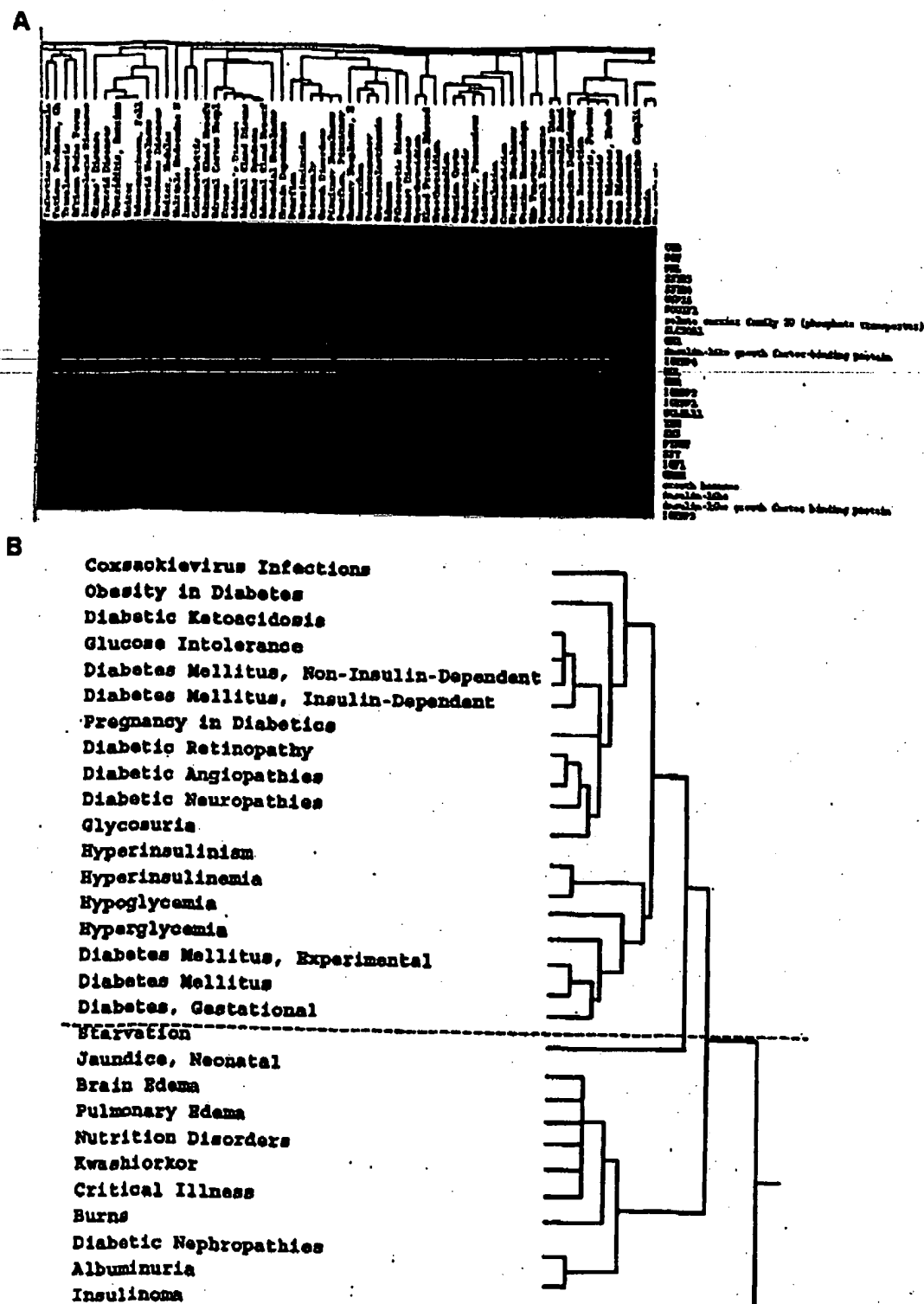


Figure 2. Global validation by clustering analysis. 2(A). The gene sets and the corresponding LPF values for 1000 diseases, each with at least 50 gene relationships, were used in an unsupervised clustering of the diseases based on the gene patterns associated with them. A sample of the data is shown here. 2(B). One of the resulting clusters is shown that corresponds to blood sugar states. Diabetes terms (above the line) and starvation states terms (under the line) clustered together. Within these groups, there is also clustering of diabetic small vessel complications, altered serum chemistries, nutritional disorders, etc. (Supplemental Figure 1: [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Figure\\_1.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Figure_1.html)).

Finally, to validate our findings, we computed similar correlations between the breast cancer expression data and LPF scores generated by MedGene for hypertension, a

disease unrelated to breast cancer. As expected, we did not observe an increasing trend in correlation for hypertension.





410 Journal of Proteome Research • Vol. 2, No. 4, 2003

Table 2. Top 25 Genes Related to Selected Human Diseases\*

breast neoplasms	hypertension	rheumatoid arthritis	bipolar disorder	atherosclerosis
estrogen receptor	<i>REN</i>	<i>RA</i>	<i>ERDA1</i>	apolipoprotein
<i>PCR</i>	<i>DBP</i>	<i>TNFRSF10A</i>	<i>SNAP29</i>	<i>APOE</i>
<i>ERBB2</i>	<i>LEP</i>	<i>CRP</i>	<i>PFKL</i>	<i>LDLR</i>
<i>BRCA1</i>	<i>AGT</i>	<i>AS</i>	<i>DRD2</i>	<i>ELN</i>
<i>BRCA2</i>	<i>INS</i>	<i>ESR1</i>	<i>TRH</i>	<i>ARG1</i>
<i>EGFR</i>	kallikrein	<i>HLA-DRB1</i>	<i>IMPA2</i>	<i>APOB</i>
<i>CYP19</i>	<i>ACE</i>	<i>DR1</i>	<i>HTR3A</i>	<i>APOA1</i>
<i>TFF1</i>	endothelin	interleukin	<i>DRD3</i>	<i>MSR1</i>
<i>PSEN2</i>	<i>S100A6</i>	<i>TNF</i>	<i>REM</i>	<i>LPL</i>
<i>TP53</i>	<i>BDK</i>	<i>IL6</i>	<i>KCNN3</i>	<i>PON1</i>
<i>CESS</i>	<i>DIAPH</i>	collagen	<i>DRD4</i>	plasminogen
<i>CEACAM5</i>	<i>SARI</i>	<i>IL1A</i>	<i>HTR2C</i>	activator inhibitor
<i>ERBB3</i>	<i>PIH</i>	<i>ACR</i>	<i>RELN</i>	<i>PLG</i>
cyclin	<i>CD59</i>	<i>TNFRSF12</i>	<i>DBH</i>	vascular cell
<i>COX5A</i>	<i>ALB</i>	<i>IL2</i>	<i>MAOA</i>	adhesion molecule
cathepsin	<i>CYP11B2</i>	<i>CHI3L1</i>	<i>COMT</i>	<i>ATOH1</i>
<i>ERBB4</i>	<i>MAT2B</i>	<i>IL8</i>	<i>HTR2A</i>	<i>VWF</i>
<i>TRAM</i>	angiotensin receptor	interleukin 1 matrix	<i>SYNJ1</i>	<i>INS</i>
<i>CCND1</i>	<i>AGTR2</i>	metalloproteinase	<i>INPP1</i>	<i>ARG2</i>
<i>EGF</i>	<i>NPPA</i>	interferon	<i>NEDD4L</i>	<i>ABCA1</i>
<i>MUC1</i>	<i>LVM</i>	<i>CD68</i>	<i>FRA13C</i>	<i>OLR1</i>
insulin-like	<i>DBH</i>	<i>IL4</i>	transducer of	collagen
<i>BCL2</i>	<i>NPY</i>	<i>IL17</i>	<i>ERBB2</i>	<i>MCP</i>
mucin	<i>POMC</i>	<i>MMP3</i>	<i>BAIAP3</i>	lipoprotein
<i>FGF3</i>	neuropeptide	<i>SIL</i>	<i>ATP1B3</i>	<i>APOA2</i>
			<i>DRD5</i>	intercellular
				adhesion molecule
				<i>RAB27A</i>

\* MedGene results for the top 25 genes associated with breast neoplasms, hypertension, rheumatoid arthritis, bipolar disorder, and atherosclerosis, respectively, ranked by LPI scores. The hyperlink to all the papers co-citing the gene and the disease is available at MedGene website (<http://hipseq.med.harvard.edu/MedGene/>).

## Discussion

The Human Genome Project heralded a new era in biological research where the emphasis on understanding specific pathways has expanded to global studies of genomic organization and biological systems. High-throughput technologies can provide novel insight into comprehensive biological function but also introduces new challenges. The utility of these technologies is limited to the ability to generate, analyze, and interpret large gene lists. MedGene, a relational database derived by mining the information in Medline, was created to address this need. MedGene users can query for a rank-ordered list of human gene-disease relationships (Table 2) for one or more diseases. Each entry is hyperlinked to the original papers supporting each association and to other relevant databases.

MedGene is an innovative extension of previous text mining approaches. Perez-Iratxeta et al. used the GO annotation and their chromosomal locations to predict genes that may contribute to inherited disorders.<sup>8</sup> MedGene takes a broader view and includes all diseases and all possible gene-disease relationships. Furthermore, MedGene utilizes co-citation to indicate a relationship rather than GO annotation, which is limited to the subset of genes that have GO annotation. Our approach is complementary to that taken by Chaussabel and Sher, who used the frequency of co-cited terms to cluster genes into a hierarchy of gene-gene relationships.<sup>9</sup>

A unique aspect of this tool is the ability to assess the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation. This presupposes that most co-citations describe a positive association, often referred to as publication bias<sup>15</sup> and is supported by our observations

that negative associations are rare (Supplemental Table 3: [http://hipseq.med.harvard.edu/MedGene/publication/s\\_Table3.html](http://hipseq.med.harvard.edu/MedGene/publication/s_Table3.html)). Of course, relationships established by frequency of co-citation do not necessarily represent a true biological link; however, it is strong evidence to support a true relationship.

Another important feature of MedGene is the implementation of software filters that substantially reduced the error rate. We estimate that less than 10% of all associations were missed and at least 70% of even the weakest associations were real. For this study, all of the filters that we applied were general ones, e.g., expanding the list of all gene names to address the different syntax forms used by different journals, eliminating gene names that correspond to common English words, etc. The majority of the remaining search-term ambiguities were idiosyncratic and difficult to identify systematically without causing a significant rise in false negatives. Alternative approaches, such as the examination of the nearest neighbor terms, need to be considered to further reduce the false positive rate.

It is not uncommon to see expression changes in microarray experiments as small as 2-fold reported in the literature. Even when these expression changes are statistically significant, it is not always clear if they are biologically meaningful. When comparing expression levels of disease to normal tissue, one expects an enrichment of known disease-related genes to appear in the altered expression group. MedGene provided a unique opportunity to test this notion in the context of existing knowledge on a novel breast cancer microarray dataset. For genes displaying a 5-fold change or less in tumors compared to normal, there was no evidence of a correlation between altered gene expression and a known role in the disease. This

Table 3. Genes with Large Expression Changes in ER- but Not in ER+ Breast Tumors

gene symbol	fold change (ER+)	fold change (ER-)
KRTHB1	1.0	610.8
BRS3	1.2	89.4
DKK1	1.2	69.8
ZIC1	1.9	59.6
TLR1	1.0	38.5
KIAA0680	2.6	33.2
CDKN3	1.0	30.6
EBI2	4.0	27.9
GZMB	3.8	21.9
STK18	4.7	18.6
GPR49	1.0	14.6
MYO10	1.6	14.4
LAD1	-1.0	13.5
POLE2	4.2	13.0
HMG4	4.4	12.9
BCL2L11	-1.2	12.3
LRP8	2.9	12.2
CCNB2	1.0	11.8
CCNE2	4.0	11.6
FCB	-4.3	11.1
KNSL6	2.9	10.9
HIF5	3.0	10.2
SERPINH2	4.6	10.2
YAP1	1.0	10.0
LPHB	-1.3	-10.4
TCEA2	-1.1	-10.8
TFF1	1.3	-11.4
COL17A1	-4.1	-15.7
POP5	1.1	-16.2
BPAG1	-4.6	-22.3
PDZK1	-1.1	-36.8
VECFC	-2.8	-51.5
MUC6	-1.4	-64.9
SERPINA5	-1.0	-83.1
MEIS1	-1.6	-85.9
CA12	2.4	-150.3

Table 3. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative, but not ER positive breast tumors. All of these genes have either never been co-cited with breast cancer or have a weak association except those marked with an \*.

reflects the many genes whose role in breast cancer may not involve large changes in expression in sporadic tumors (e.g., *BRCA1* and *BRCA2*) and genes whose modest changes in expression may be unrelated to the disease. Strikingly, among genes with a 10-fold change or more in expression level, there was a strong and significant correlation between expression level and a published role in the disease, providing the first global validation of the micro-array approach to identifying disease-specific genes.

The results derived from MedGene have two implications. First, a careful hunt for corroborating evidence of a role in breast cancer should precede any further study of genes with less than 5-fold expression level changes. Second, any genes with 10-fold changes or more are likely to be related to breast cancer and warrant attention. It is likely that this threshold will change depending on the disease as well as the experiment.

Interestingly, the observed correlation was only found among ER-positive tumors, not ER-negative. This may reflect a bias in the literature to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently. The MedGene approach identified a set of relatively understudied, yet highly expressed genes in ER-negative tumors that are worthy of further examination (Table 3).

In conclusion, we have developed an automated method of summarizing and organizing the vast biomedical literature. To our knowledge, the resulting database is the most comprehensive and accurate of its kind. By generating a score that reflects the strength of the association, it provides an important tool for the rapid and flexible analysis of large datasets from various high-throughput screening experiments. Furthermore, it can be used for selecting subsets of genes for functional studies, for building disease-specific arrays, for looking at genes common to multiple diseases and various other high-throughput applications. In the future, it will be possible to enhance the utility of the MedGene database by building links between genes and other MeSH terms as well as other biological processes and concepts, such as cell division and responses to small molecules.

**Acknowledgment.** We would like to thank P. Braun, L. Garraway, J. Pearlberg, and other members of our institute for helpful discussion. Many thanks to the NLM (National Library of Medicine) for licensing of MEDLINE and the annotation effort of adding MeSH indexes for MEDLINE abstracts. This work was funded by grants from the Breast Cancer Research Foundation and an NHLBI PGA Grant (Vol HL66582-02).

**Supporting Information Available:** Twenty-three human disease category headings along with all of their child terms selected from the 2002 MeSH Index (Supplemental Table 1); analysis of the causes of false negatives in MedGene (Supplemental Table 2); meaningful gene-disease relationships found in MedGene (Supplemental Table 3); causes for incorrect assignment of gene indexes (Supplemental Table 4); a review of the results, showing that the resulting disease clusters were indeed logical (Supplemental Figure 1); and a review of the results showing that among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease (Supplemental Figure 2). This material is available free of charge via the Internet at <http://pubs.acs.org> and at the web sites mentioned in the text.

## References

- (1) Baasiri, R. A.; Glasser, S. R.; Steffen, D. L.; Wheeler, D. A. *Oncogene* 1999, 18, 7958-7965.
- (2) Steffen, D. L.; Levine, A. E.; Yarus, S.; Baasiri, R. A.; Wheeler, D. A. *Bioinformatics* 2000, 16, 639-649.
- (3) Marcotte, E. M.; Xenarios, I.; Eisenberg, D. *Bioinformatics* 2001, 17, 359-363.
- (4) Ono, T.; Hishigaki, H.; Tanigami, A.; Takagi, T. *Bioinformatics* 2001, 17, 155-161.
- (5) Jensen, T. K.; Laegreid, A.; Komorowski, J.; Hovig, E. *Nat. Genet.* 2001, 28, 21-28.
- (6) Chaussabel, D.; Sher, A. *Genome Biol.* 2002, 3, RESEARCH0055.
- (7) Gibbons, F. D.; Roth, F. P. *Genome Res.* 2002, 12, 1574-1581.
- (8) Perez-Iratxeta, C.; Bork, P.; Andrade, M. A. *Nat. Genet.* 2002, 31, 316-319.
- (9) Funk, M. E.; Reid, C. A. *Bull. Med. Lib. Assoc.* 1983, 71, 176-183.
- (10) Humphrey, S. M.; Miller, N. E. *J. Am. Soc. Inf. Sci.* 1987, 38, 184-196.
- (11) Maglott, D. R.; Katz, K. S.; Scoute, H.; Pruitt, K. D. *Nucleic Acids Res.* 2000, 28, 126-128.
- (12) Pruitt, K. D.; Maglott, D. R. *Nucleic Acids Res.* 2001, 29, 137-140.
- (13) Wadelius, M.; Andersson, A. O.; Johansson, J. E.; Wadelius, C.; Rane, E. *Pharmacogenetics* 1999, 9, 333-340.
- (14) Adam, R. M.; Borer, J. G.; Williams, J.; Eastham, J. A.; Loughlin, K. R.; Freeman, M. R. *Endocrinology* 1999, 140, 5866-5875.
- (15) Montori, V. M.; Smieja, M.; Guyatt, G. H. *Mayo Clin. Proc.* 2000, 75, 1284-1288.
- (16) Denenberg, V. H. *Statistics Experimental Design for Behavioral and Biological Researchers*; Wiley-Liss: New York, 1976.
- (17) Rebhan, M.; Chalifa-Caspi, V.; Prilusky, J.; Lancet, D. *Trends Genet.* 1997, 13, 163.
- (18) Baloch, A.; Apweiler, R. *Nucleic Acids Res.* 2000, 28, 45-48. PR0340227

# Aneuploidy and cancer

Subrata Sen, PhD

Numeric aberrations in chromosomes, referred to as aneuploidy, is commonly observed in human cancer. Whether aneuploidy is a cause or consequence of cancer has long been debated. Three lines of evidence now make a compelling case for aneuploidy being a discrete chromosome mutation event that contributes to malignant transformation and progression process. First, precise assay of chromosome aneuploidy in several primary tumors with *in situ* hybridization and comparative genomic hybridization techniques have revealed that specific chromosome aneusomies correlate with distinct tumor phenotypes. Second, aneuploid tumor cell lines and *in vitro* transformed rodent cells have been reported to display an elevated rate of chromosome instability, thereby indicating that aneuploidy is a dynamic chromosome mutation event associated with transformation of cells. Third, and most important, a number of mitotic genes regulating chromosome segregation have been found mutated in human cancer cells, implicating such mutations in induction of aneuploidy in tumors. Some of these gene mutations, possibly allowing unequal segregations of chromosomes, also cause tumorigenic transformation of cells *in vitro*. In this review, the recent publications investigating aneuploidy in human cancers, rate of chromosome instability in aneuploidy tumor cells, and genes implicated in regulating chromosome segregation found mutated in cancer cells are discussed. Curr Opin Oncol 2000, 12:82-88 © 2000 Lippincott Williams & Wilkins, Inc.

The University of Texas, M.D. Anderson Cancer Center, Department of Laboratory Medicine, Houston, Texas, USA

Correspondence to Subrata Sen, PhD, The University of Texas, MD Anderson Cancer Center, Department of Laboratory Medicine, Box 054, 1515 Holcombe Blvd., Houston, TX 77030, USA; tel: 713-792-2560; fax: 713-792-4094; e-mail: ssen@mdanderson.org

Current Opinion in Oncology 2000 12:82-88

## Abbreviations

CGH comparative genomic hybridization  
CHE Chinese hamster embryo cells  
FISH fluorescence in situ hybridization  
HPRC hereditary papillary renal carcinoma  
ISH *in situ* hybridization

ISSN 1040-8746 © 2000 Lippincott Williams & Wilkins, Inc.

Cancer research over the past decade has firmly established that malignant cells accumulate a large number of genetic mutations that affect differentiation, proliferation, and cell death processes. In addition, it is also recognized that most cancers are clonal, although they display extensive heterogeneity with respect to karyotypes and phenotypes of individual clonal populations. It is estimated that numeric chromosomal imbalance, referred to as *aneuploidy*, is the most prevalent genetic change recorded among over 20,000 solid tumors analyzed thus far [1]. Phenotypic diversity of the clonal populations in individual tumors involve differences in morphology, proliferative properties, antigen expression, drug sensitivity, and metastatic potentials. It has been proposed that an underlying acquired genetic instability is responsible for the multiple mutations detected in cancer cells that lead to tumor heterogeneity and progression [2]. In a somewhat contradictory argument, it has also been suggested that clonal expansion due to selection of cells undergoing normal rates of mutation can explain malignant transformation and progression process in humans [3]. Acquired genetic instability, nonetheless, is considered important for more rapid progression of the disease [4••]. Although the original hypothesis on genetic instability in cancer primarily focused on chromosome imbalances in the form of aneuploidy in tumor cells, the actual relevance of such mutations in cancer remains a controversial issue.

Whether or not aneuploidy contributes to the malignant transformation and progression process has long been debated. A prevalent idea on genetics of cancer referred to as "somatic gene mutation hypothesis" contends that gene mutations at the nucleotide level alone can cause cancer by either activating cellular proto-oncogenes to dominant cancer causing oncogenes and/or by inactivating growth inhibitory tumor suppressor genes. In this scheme of things chromosomal instability in the form of aneuploidy is a mere consequence rather than a cause of malignant transformation and progression process.

In this review, some of the recent observations on the subject are discussed and compelling evidence is provided to suggest that aneuploidy is a distinct form of genetic instability in cancer that frequently correlates with specific phenotypes and stages of the disease. Furthermore, discrete genetic targets affecting chromosomal stability in cancer cells, recently identified, are also discussed. These data provide a new direction toward elucidating the molecular mechanisms responsi-

ble for induction of aneuploidy in cancer and may eventually be exploited as novel therapeutic targets in the future.

### Genetic alterations in cancer

Alterations in many genetic loci regulating growth, senescence, and apoptosis, identified in tumor cells, have led to the current understanding of cancer as a genetic disease. The genetic changes identified in tumors include: subtle mutations in genes at the nucleotide level; chromosomal translocations leading to structural rearrangements in genes; and numeric changes in either partial segments of chromosomes or whole chromosomes (aneuploidy) causing imbalance in gene dosage.

For the purpose of this review, both segmental and whole chromosome imbalances leading to altered DNA dosage in cancer cells are included as examples of aneuploidy.

### Incidence of aneuploidy in cancer

Evidence of aneuploidy involving one or more chromosomes have been commonly reported in human tumors. Although these observations were initially made using classic cytogenetic techniques late in a tumor's evolution and were difficult to correlate with cancer progression, more recent studies have reported association of specific nonrandom chromosome aneuploidy with different biologic properties such as loss of hormone dependence and metastatic potential [5].

Classic cytogenetic studies performed on tumor cells had serious limitations in scope because they were applicable only to those cases in which mitotic chromosomes could be obtained. Because of low spontaneous rates of cell division in primary tumors, analyses depended on cells either derived selectively from advanced metastases or those grown *in vitro* for variable periods of time. In both instances, metaphases analyzed represented only a subset of primary tumor cell population. Two major advances in cytogenetic analytic techniques, *in situ* hybridization (ISH) and comparative genomic hybridization (CGH), have allowed better resolution of chromosomal aberrations in freshly isolated tumor cells [6]. ISH analyses with chromosome-specific DNA probes, a powerful adjunct to metaphasic analysis, allows assessment of chromosomal anomalies within tumor cell populations in the contexts of whole nuclear architecture and tissue organization. CGH allows genome wide screening of chromosomal anomalies without the use of specific probes even in the absence of prior knowledge of chromosomes involved. Although both techniques have certain limitations in terms of their resolution power, they nonetheless provide a better approximation of chromosomal changes occurring among tumors of various histology, grade, and stage

compared with what was possible with the classic cytogenetic techniques. Genomic ploidy measurements have also been performed at the DNA level with flow cytometry and cytofluorometric methods. Although these assays underestimate chromosome ploidy due to a chromosomal gain occasionally masking a chromosomal loss in the same cell, several studies using these methods have supported the conclusion that DNA aneuploidy closely associates with poor prognosis in various cancers [7,8]. This discussion of some recent examples published on aneuploidy in cancer includes discussion of studies dealing with DNA ploidy measurements as well. Most of these observations are correlative without direct proof of specific involvement of genes on the respective chromosomes. Identification of putative oncogenes and tumor suppressor genes on gained and lost chromosomes in aneuploid tumors, however, are providing strong evidence that chromosomes involved in aneuploidy play a critical role in the tumorigenic process.

In renal tumors, either segmental or whole chromosome aneuploidy appears to be uniquely associated with specific histologic subtypes [9]. Tumors from patients with hereditary papillary renal carcinomas (HPRC) commonly show trisomy of chromosome 7, when analyzed by CGH. Germline mutations of a putative oncogene *MET* have been detected in patients with HPRC. A recent study [10] has demonstrated that an extra copy of chromosome 7 results in nonrandom duplication of the mutant *MET* allele in HPRC, thereby implicating this trisomy in tumorigenesis. The study suggested that mutation of *MET* may render the cells more susceptible to errors in chromosome replication, and that clonal expansion of cells harboring duplicated chromosome 7 reflects their proliferative advantage. In addition to chromosome 7, trisomy of chromosome 17 in papillary tumors and also of chromosome 8 in mesoblastic nephroma are commonly seen. Association of specific chromosome imbalances with benign and malignant forms of papillary renal tumors, therefore, not only contribute to an understanding of tumor origins and evolution, but also implicate aneuploidy of the respective chromosomes in the tumorigenic transformation process.

In colorectal tumors, chromosome aneuploidy is a common occurrence. In fact, molecular allelotyping studies have suggested that limited karyotyping data available from these tumors actually underestimate the true extent of these changes. Losses of heterozygosity reflecting loss of the maternal or paternal allele in tumors are widespread and often accompanied by a gain of the opposite allele. Therefore, for example, a tumor could lose a maternal chromosome while duplicating the same paternal chromosome, leaving the tumor cell

with a normal karyotype and ploidy but an aberrant allelotype. It has been estimated that cancer of the colon, breast, pancreas, or prostate may lose an average of 25% of its alleles. It is not unusual to discover that a tumor has lost over half of its alleles [4]. In clinical settings, DNA ploidy measurements have revealed that DNA aneuploidy indicates high risk of developing severe premalignant changes in patients with ulcerative colitis, who are known to have an increased risk of developing colorectal cancer [11]. DNA aneuploidy has been found to be one of the useful indicators of lymph node metastasis in patients with gastric carcinoma and associated with poor outcome compared with diploid cases [12,13]. CGH analyses of chromosome aneuploidy, on the other hand, was reported to correlate gain of chromosome 20q with high tumor S phase fractions and loss of 4q with low tumor apoptotic indices [14]. Aneuploidy of chromosome 4 in metastatic colorectal cancer has recently been confirmed in studies that used unbiased DNA fingerprinting with arbitrarily primed polymerase chain reactions to detect moderate gains and losses of specific chromosomal DNA sequences [15]. The molecular karyotype (amplotype) generated from colorectal cancer revealed that moderate gains of sequences from chromosomes 8 and 13 occurred in most tumors, suggesting that overrepresentation of these chromosomal regions is a critical step for metastatic colorectal cancer.

In addition to being implicated in tumorigenesis and correlated with distinct tumor phenotypes, chromosome aneuploidy has been used as a marker of risk assessment and prognosis in several other cancers. The potential value of aneuploidy as a noninvasive tool to identify individuals at high risk of developing head and neck cancer appears especially promising. Interphase fluorescence *in situ* hybridization (FISH) revealed extensive aneuploidy in tumors from patients with head and neck squamous cell carcinomas (HNSCC) and also in clinically normal distant oral regions from the same individuals [16,17]. It has been proposed that a panel of chromosome probes in FISH analyses may serve as an important tool to detect subclinical tumorigenesis and for diagnosis of residual disease. The presence of aneuploid or tetraploid populations is seen in 90% to 95% of esophageal adenocarcinomas, and when seen in conjunction with Barrett's esophagus, a premalignant condition, predicts progression of disease [18,19]. Chromosome ploidy analyses in conjunction with loss of heterozygosity and gene mutation studies in Barrett's esophagus reflect evolution of neoplastic cell lineages *in vivo* [20]. Evolution of neoplastic progeny from Barrett's esophagus following somatic genetic mutations frequently involves bifurcations and loss of heterozygosity at several chromosomal loci leading to aneuploidy and cancer. Accordingly, it is hypothesized that during

tumor cell evolution diploid cell progenitors with somatic genetic abnormalities undergo expansion with acquired genetic instability. Such instability, often manifested in the form of increased incidence of aneuploidy, enters a phase of clonal evolution beginning in premalignant cells that proceeds over a period of time and occasionally leads to malignant transformation. The clonal evolution continues even after the emergence of cancer.

The significance of DNA and chromosome aneuploidy in other human cancers continue to be evaluated. Among papillary thyroid carcinomas, aneuploid DNA content in tumor cells was reported to correlate with distant metastases, reflecting worsened prognosis [21]. Genome wide screening of follicular thyroid tumors by CGH, on the other hand, revealed frequent loss of chromosome 22 in widely invasive follicular carcinomas [22]. Chromosome copy number gains in invasive neoplasm compared with foci of ductal carcinoma *in situ* (DCIS) with similar histology have been proposed to indicate involvement of aneuploidy in progression of human breast cancer [23]. ISH analyses of cervical intraepithelial neoplasia has provided suggestive evidence that chromosomes 1, 7 and X aneusomy is associated with progression toward cervical carcinoma [24].

Although the prognostic value of numeric aberrations remains a matter of debate in human hematopoietic neoplasia, there have been recent studies to suggest that the presence of monosomy 7 defines a distinct subgroup of acute myeloid leukemia patients [25]. It is interesting in this context that therapy-related myelodysplastic syndromes have been reported to display monosomy 5 and 7 karyotypes, reflecting poor prognosis [26].

The clinical observations, mentioned previously, are supported by *in vitro* studies in human and rodent cells in which aneuploidy is induced at early stages of transformation [27,28]. It is even suggested that aneuploidy may cause cell immortalization, in some instances, that is a critical step preceeding transformation.

Finally, in an interesting study to develop transgenic mouse models of human chromosomal diseases, chromosome segment specific duplication and deletions of the genome were reported to be constructed in mouse embryonic stem cells [29]. Three duplications for a portion of mouse chromosome 11 syntenic with human chromosome 17 were established in the mouse germline. Mice with 1Mb duplication developed corneal hyperplasia and thymic tumors. The findings represent the first transgenic mouse model of aneuploidy of a defined chromosome segment that documents the direct role of chromosome aneusomy in tumorigenesis.

### **Aneuploidy as "dynamic cancer-causing mutation" instead of a "consequential state" in cancer**

According to the hypothesis previously discussed, aneuploidy represents either a "gain of function" or "loss of function" mutation at the chromosome level with a causative influence on the tumorigenesis process. The hypothesis, however, is based only on circumstantial evidence even though existence of aneuploidy is correlated with different tumor phenotypes. The existence of numeric chromosomal alterations in a tumor does not mean that the change arose as a dynamic mutation due to genomic instability, because several factors could lead to consequential aneuploidy in tumors, also. Although aneuploidy as a dynamic mutation due to genomic instability in tumor cells would occur at a certain measurable rate per cell generation, a consequential state of aneuploidy in tumors may not occur at a predictable rate under similar conditions or in tumors with similar phenotypes. In addition to genomic instability, differences in environmental factors with selective pressure, could explain high incidence of aneuploidy and other somatic mutations in tumors compared with normal cells [4]. These include humoral, cell substratum, and cell-cell interaction differences between tumor and normal cell environments. It could be argued that despite similar rates of spontaneous aneuploidy induction in normal and tumor cells, the latter are selected to proliferate due to altered selective pressure in the tumor cell environment, whereas the normal cells are eliminated through activation of apoptosis. Alternatively, of course, one could postulate that selective expression or overexpression of anti-apoptotic proteins or inactivation of proapoptotic proteins in tumor cells may counteract default induction of apoptosis in G2/M phase cells undergoing missegregation of chromosomes. Recent demonstration of overexpression of a G2/M phase anti-apoptotic protein survivin in cancer cells [30] suggests that this protein may favor aberrant progression of aneuploid transformed cells through mitosis. This would then lead to proliferation of aneuploid cell lineages, which may undergo clonal evolution.

To ascertain that aneuploidy is a dynamic mutational event, various human tumor cell lines and transformed rodent cell lines have been analyzed for the rate of aneuploidy induction. When grown under controlled *in vitro* conditions, such conditions ensure that environmental factors do not influence selective proliferation of cells with chromosome instability. In one study, Lengauer *et al.* [31•] provided unequivocal evidence by FISH analyses that losses or gains of multiple chromosomes occurred in excess of  $10^{-2}$  per chromosome per generation in aneuploid colorectal cancer cell lines. The study further concluded that such chromosomal instability appeared to be a dominant trait. Using another *in*

*vitro* model system of Chinese hamster embryo (CHE) cells, Duesberg *et al.* [32•] have also obtained similar results. With clonal cultures of CHE cells, transformed with nongenotoxic chemicals and a mitotic inhibitor, these authors demonstrated that the overwhelming majority of the transformed colonies contained more than 50% aneuploid cells, indicating that aneuploidy would have originated from the same cells that underwent transformation. All the transformed colonies tested were tumorigenic. It was further documented that the ploidy factor representing the quotient of the modal chromosome number divided by the normal diploid number, in each clone, correlated directly with the degree of chromosomal instability. Therefore, chromosomal instability was found proportional to the degree of aneuploidy in the transformed cells and the authors hypothesized that aneuploidy is a unique mechanism of simultaneously altering and destabilizing, in a massive manner, the normal cellular phenotypes. In the absence of any evidence that the transforming chemicals used in the study did not induce other somatic mutations, it is difficult to rule out the contribution of such mutations in the transformation process. These results nonetheless make a strong case for aneuploidy being a dynamic chromosome mutation event intimately associated with cancer.

### **Aneuploidy versus somatic gene mutation in cancer**

The idea that numeric chromosome imbalance or aneuploidy is a direct cause of cancer was proposed at the turn of the century by Theodore Boveri [33]. However, the hypothesis was largely ignored over the last several decades in favor of the somatic gene mutation hypothesis, mentioned earlier. Evidence accumulating in the literature lately on specific chromosome aneusomies recognized in primary tumors, incidence of aneuploidy in cells undergoing transformation, and aneuploid tumor cells showing a high rate of chromosome instability have led to the rejuvenation of Boveri's hypothesis. The concept has recently been discussed as a "vintage wine in a new bottle" [34•]. The author points out that except for rare cancers caused by dominant retroviral oncogenes, diploidy does not seem to occur in solid tumors, whereas aneuploidy is a rule rather than exception in cancer.

Aneuploidy as an effective mutagenic mechanism driving tumor progression, on the other hand, is being recognized as a viable solution to the paradox that with known mutation rate in non-germline cells ( $\sim 10^{-7}$  per gene per cell generation) tumor cell lineages cannot accumulate enough mutant genes during a human lifetime [35]. The concept is gaining significant credibility since genes that potentially affect chromosome segregation were found mutated in human cancer. Some of

these genes have also been shown to have transforming capability in *in vitro* assays. Selected recent publications describing the findings are being discussed below in reference to the mitotic targets potentially involved in inducing chromosome segregation anomalies in cells.

### Potential mitotic targets and molecular mechanisms of aneuploidy

Because aneuploidy represents numeric imbalance in chromosomes, it is reasonable to expect that aneuploidy arises due to missegregation of chromosomes during cell division. There are many potential mitotic targets, which could cause unequal segregation of chromosomes (Fig. 1). Recent investigations have identified several genes involved in regulating these mitotic targets and mitotic checkpoint functions, which can be implicated in induction of aneuploidy in tumor cells. This discussion is restricted to those mitotic targets and checkpoint genes whose abnormal functioning has been observed in cancer or has been shown to cause tumorigenic transformation of cells, in recent years. The role of telomeres is discussed elsewhere in this issue. For a more detailed description of the components of mitotic machinery and their possible involvement in causing chromosome segregation abnormalities in tumor cells, readers may refer to a recently published review [36•].

Among the mitotic targets implicated in cancer, centrosome defects have been observed in a wide variety of malignant human tumors. Centrosomes play a central role in organizing the microtubule network in interphase cells and mitotic spindle during cell division. Multipolar mitotic spindles have been observed in human cancers *in situ* and abnormalities in the form of supernumerary

centrosomes, centrosomes of aberrant size and shape as well as aberrant phosphorylation of centrosome proteins have been reported in prostate, colon, brain, and breast tumors [37,38]. In view of the findings that abnormal centrosomes retain the ability to nucleate microtubules *in vitro*, it is conceivable that cells with abnormal centrosomes may missegregate chromosomes producing aneuploid cells. The molecular and genetic bases of abnormal centrosome generation and the precise pathway through which they regulate the chromosome segregation process remain to be elucidated. Recent discovery of a centrosome-associated kinase STK15/BTAK/aurora2, naturally amplified and overexpressed in human cancers, has raised the interesting possibility that aberrant expression of this kinase is critically involved in abnormal centrosome function and unequal chromosome segregation in tumor cells [39,40]. Exogenous expression of the kinase in rodent and human cells was found to correlate with an abnormal number of centrosomes, unequal partitioning of chromosomes during division, and tumorigenic transformation of cells. It is relevant in this context to mention that the *Xenopus* homologue of human STK15/BTAK/aurora2 kinase has recently been shown to phosphorylate a microtubule motor protein XIEg5, the human orthologue of which is known to participate in the centrosome separation during mitosis [41]. Findings on STK15/aurora2 kinase, thus, provide an interesting lead to a possible molecular mechanism of centrosome's role in oncogenesis. Centrosomes have, of late, been implicated in oncogenesis from studies revealing supernumerary centrosomes in *p53*-deficient fibroblasts and overexpression of another centrosome kinase PLK1 being detected in human non-small cell lung cancer [42].

One of the critical events that ensures equal partitioning of the chromosomes during mitosis is the proper and timely separation of sister chromatids that are attached to each other and to the mitotic spindle. Untimely separation of sister chromatids has been suspected as a cause of aneuploidy in human tumors. Cohesion between sister chromatids is established during replication of chromosomes and is retained until the next metaphase/anaphase transition. It has been shown that during metaphase-anaphase transition, the anaphase promoting complex/cyclosome triggers the degradation of a group of proteins called securins that inhibit sister chromatid separation. A vertebrate securin (v-securin) has recently been identified that inhibits sister chromatid separation and is involved in transformation and tumorigenesis. Subsequent analysis revealed that the human securin is identical to the product of the gene called pituitary tumor transforming gene, which is overexpressed in some tumors and exhibits transforming activity in NIH3T3 cells. It is proposed that elevated expression of the v-securin may contribute to generation of malignant tumors due to

**Figure 1. Potential mitotic targets causing aneuploidy in oncogenesis**

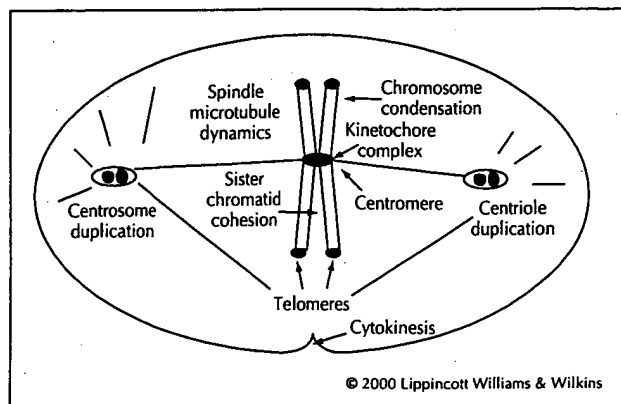


Diagram illustrates that defects in several processes involving chromosomal, spindle microtubule, and centrosomal targets, in addition to abnormal cytokinesis, may cause unequal partitioning of chromosomes during mitosis, leading to aneuploidy. Recently obtained evidence in favor of some of these possibilities is discussed in the text.



chromosome gain or loss produced by errors in chromatid separation [43•].

Normal progression through mitosis during prophase to anaphase transition is monitored at least at two checkpoints. One checkpoint operates during early prophase at G2 to metaphase progression while the second ensures proper segregation of chromosomes during metaphase to anaphase transition. Several mitotic checkpoint genes responding to mitotic spindle defects have been identified in yeast. The metaphase-anaphase transition is delayed following activation of this checkpoint during which kinetochores remain unattached to the spindle. The signal is transmitted through a kinetochore protein complex consisting of Mps1p and several Mad and Bub proteins [44]. It is expected that for unequal chromosome segregation to be perpetuated through cell proliferation cycles giving rise to aneuploidy, checkpoint controls have to be abrogated.

Following this logic, Vogelstein *et al.* [45•] hypothesized that aneuploid tumors would reveal mutation in mitotic spindle checkpoint genes. Subsequent studies by these investigators have proven the validity of this hypothesis and a small fraction of human colorectal cancers have revealed the presence of mutations in either hBub1 or hBubR1 checkpoint genes. It was further revealed that mutant BUB1 could function in a dominant negative manner conferring an abnormal spindle checkpoint when expressed exogenously. Inactivation of spindle checkpoint function in virally induced leukemia has also recently been documented following the finding that hMAD1 checkpoint protein is targeted by the Tax protein of the human T-cell leukemia virus type 1. Abrogation of hMAD1 function leads to multinucleation and aneuploidy [46].

In addition to mitotic spindle checkpoint defects, failed DNA damage checkpoint function in yeast is frequently associated with aberrant chromosome segregation as well. It, therefore, appears intriguing yet relevant that the human *BRCA1* gene, proposed to be involved in DNA damage checkpoint function, when mutated by a targeted deletion of exon 11 led to defective G2/M cell cycle checkpoint function and genetic instability in mouse embryonic fibroblasts [47]. The cells revealed multiple functional centrosomes and unequal chromosome segregation and aneuploidy. Although the molecular basis for these abnormalities is not known at this time, it raises the interesting possibility that such an aneuploidy-driven mechanism may be involved in tumorigenesis in individuals carrying germline mutations of *BRCA1* gene.

## Conclusion

Growing evidence from human tumor cytogenetic investigations strongly suggest that aneuploidy is associated with the development of tumor phenotypes. Clinical findings of correlation between aneuploidy and tumorigenesis are supported by studies with *in vitro* grown transformed cell lines. Molecular genetic analyses of tumor cells provide credible evidence that mutations in genes controlling chromosome segregation during mitosis play a critical role in causing chromosome instability leading to aneuploidy in cancer. Further elucidation of molecular and physiologic bases of chromosome instability and aneuploidy induction could lead to the development of new therapeutic approaches for common forms of cancer.

## Acknowledgments

The author is thankful to Drs. Bill Brinkley and Pramila Sen for discussions and advice. Help from Ms. Donna Sprabary and Ms. Hongyi Zhou in preparation of this manuscript is gratefully acknowledged. The work in the author's laboratory was supported by grants from the NIH and The University of Texas M.D. Anderson Cancer Center.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- Of special interest
- Of outstanding interest

- 1 Heim S, Mitelman F: Cancer cytogenetics, edn 2. New York: Wiley Liss Inc., 1995.
- 2 Nowell PC: The clonal evolution of tumor cell populations. *Science* 1976, 194:23-28.
- 3 Tomlinson IP, Novelli MR, Bodmer WF: The mutation rate and cancer. *Proc Natl Acad Sci USA* 1996, 93:14800-14803.
- 4 Lengauer C, Kinzler KW, Vogelstein B: Genetic instabilities in human cancers. *Nature* 1998, 396:643-649.
- An excellent review on the significance and possible mechanisms of genetic instability in cancer.
- 5 Heppner GH, Miller FR: The cellular basis of tumor progression. *Int Rev Cytol* 1998, 177:1-56.
- 6 Wolman SR: Chromosomal markers: signposts on the road to understanding neoplastic disease. *Diag Cytopath* 1998, 18:18-23.
- 7 Ross JS: DNA ploidy and cell cycle analysis in cancer diagnosis and prognosis. *Oncology* 1996, 10:867-890.
- 8 Magennis DP: Nuclear DNA in histological and cytological specimens: measurement and prognostic significance. *Br J Biomed Sci* 1997, 54:140-148.
- 9 Fletcher JA: Renal and bladder cancers. In: *Human Cytogenetic Cancer Markers*. Edited by Wolman SR, Sell S. Totowa, NJ: Humana Press; 1997:169-202.
- 10 Zhuang Z, Park WS, Pack S, Schmidt L, Vortmeyer AO, Pak E, et al.: Trisomy 7-harboring non-random duplication of the mutant MET allele in hereditary papillary renal carcinomas. *Nat Genet* 1998, 20:66-69.
- 11 Lindberg JO, Stenling RB, Rutegard JN: DNA aneuploidy as a marker of premalignancy in surveillance of patients with ulcerative colitis. *Br J Surg* 1999, 86:947-950.
- 12 Sasaki O, Kido K, Nagahama S: DNA ploidy, Ki-67 and p53 as indicators of lymph node metastasis in early gastric carcinoma. *Anal Quant Cytol Histol* 1999, 21:85-88.
- 13 Abad M, Ciudad J, Rincon MR, Silva I, Paz-Bouza JI, Lopez A, et al.: DNA aneuploidy by flow cytometry is an independent prognostic factor in gastric cancer. *Anal Cell Path* 1998, 16:223-231.
- 14 DeAngelis PM, Clausen OP, Schjolberg A, Stokke T: Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their

- associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 1999, 80:526-535.
- 15 Malkhosyan S, Yasuda J, Scotto JL, Sekiya T, Yokota J, Perucho M: Molecular karyotype (amplotype) of metastatic colorectal cancer by unbiased arbitrarily primed PCR DNA fingerprinting. *Proc Natl Acad Sci (USA)* 1998, 95:10170-10175.
  - 16 Ai H, Barrera JE, Pan Z, Meyers AD, Varella-Garcia M: Identification of individuals at high risk for head and neck carcinogenesis using chromosome aneuploidy detected by fluorescence in situ hybridization. *Mut Res* 1999, 439:223-232.
  - 17 Barrera JE, Ai H, Pan Z, Meyers AD, Varella-Garcia M: Malignancy detection by molecular cytogenetics in clinically normal mucosa adjacent to head and neck tumors. *Arch Otolaryngol Head Neck Surg* 1998, 124:847-851.
  - 18 Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, Levine DS, et al.: 17p (p53) allelic loss, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's oesophagus. *Proc Natl Acad Sci USA* 1996, 93:7081-7084.
  - 19 Teodori L, Gohde W, Persiani M, Ferrario F, Tirindelli Danesi D, Scarpignato C, et al.: DNA/protein flow cytometry as a predictive marker of malignancy in dysplasia-free Barrett's esophagus: thirteen-year follow up study on a cohort of patients. *Cytometry* 1998, 34:257-263.
  - 20 Barrett MT, Sanchez CA, Prevo LJ, Wong DJ, Galipeau PC, Paulson TG, et al.: Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat Genet* 1999, 22:106-109.
  - 21 Sturgis CD, Caraway NP, Johnston DA, Sherman SI, Kidd L, Katz RL: Image analysis of papillary thyroid carcinoma fine needle aspirates: significant association between aneuploidy and death from disease. *Cancer* 1999, 87:155-160.
  - 22 Hemmer S, Wasenius VM, Knuutila S, Joensuu H, Franssila K: Comparison of benign and malignant follicular thyroid tumours by comparative genomic hybridization. *Br J Cancer* 1998, 78:1012-1017.
  - 23 Mendelin J, Grayson M, Wallis T, Visscher DW: Analysis of chromosome aneuploidy in breast carcinoma progression by using fluorescence in situ hybridization. *Lab Inv* 1999, 79:387-393.
  - 24 Bulten J, Poddighe PJ, Robben JC, Gemmink JH, deWilde PC, Hanselaar GAGJM: Interphase cytogenetic analysis of cervical intraepithelial neoplasia. *Am J Pathol* 1998, 152:495-503.
  - 25 Krauter J, Ganzer A, Bergmann L, Raghavachar A, Hoelzer D, Lübbert M, et al.: Association between structural and numerical chromosomal aberrations in acute myeloblastic leukemia: a study by RT-PCR and FISH in 447 patients with de novo AML. *Ann Hematol* 1999, 78:265-269.
  - 26 Van Den Neste E, Louviaux I, Michaux JL, Delannoy A, Michaux L, Hagemeijer A, et al.: Myelodysplastic syndrome with monosomy 5 and/or 7 following therapy with 2-chloro-2'-deoxyadenosine. *Br J Hematol* 1999, 105:268-270.
  - 27 Namba M, Mihara K, Fushimi K: immortalization of human cells and its mechanisms. *Crit Rev Oncog* 1996, 7:19-31.
  - 28 Li R, Yerganian G, Duesberg P, Kraemer A, Willer A, Rausch C, Hehlmann R: Aneuploidy correlated 100% with chemical transformation of Chinese hamster cells. *Proc Natl Acad Sci USA* 1997, 94:14506-14511.
  - 29 Liu P, Zhang H, McLellan A, Vogel H, Bradley A: Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11. *Genetics* 1998, 150:1155-1168.
  - 30 Li F, Ambrosini G, Chu EY, Plescia J, Tognin S, Marchisio PC, Altieri DC: Control of apoptosis and mitotic spindle checkpoint survivin. *Nature* 1998, 396:580-584.
  - 31 Lengauer C, Kinzler KW, Vogelstein B: Genetic instability in colorectal cancers. *Nature* 1997, 386:623-627.  
Demonstrates chromosomal instability in aneuploid colorectal tumor cells.
  - 32 Duesberg P, Rausch C, Rasnick D, Hehlmann R: Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc Natl Acad Sci USA* 1998, 95:13692-13697.  
Correlates aneuploidy and transformation in *in vitro* grown CHE cells.
  - 33 Boveri T: Zur Frage der Entstehung maligner Tumoren. Jena, Verlag von Gustav Fischer, 1914.
  - 34 Bialy H: Aneuploidy and cancer: vintage wine in a new bottle? *Nat Biotech* 1998, 16:137-138.  
Discusses the significance of aneuploidy and gene mutations in cancer.
  - 35 Orr-Weaver TL, Weinberg RA: A checkpoint on the road to cancer. *Nature* 1998, 392:223-224.
  - 36 Pihan GA, Duxsey SJ: The mitotic machinery as a source of genetic instability in cancer. *Semin Cancer Biol* 1999, 9:289-302.  
Describes various components and regulatory mechanisms of mitotic machinery and possible mechanisms of chromosome missegregation in cancer.
  - 37 Pihan GA, Purohit A, Wallace J, Knecht H, Udda B, Queensberry P, Duxsey SJ: Centrosome defects and genetic instability in malignant tumors. *Cancer Res* 1998, 58:3974-3985.
  - 38 Lingle WL, Lutz WH, Ingle JN, Maihle NJ, Salisbury JL: Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. *Proc Natl Acad Sci USA* 1998, 95:2950-2955.
  - 39 Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, et al.: Tumor amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet* 1998, 20:189-193.  
Describes oncogenic property of centrosome associated STK15/aurora2 kinase and its involvement in aneuploidy induction.
  - 40 Bischoff JR, Anderson L, Shu Y, Morsie K, Ng I, Chan CS, et al.: A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* 1998, 17:3052-3065.  
Describes oncogenic property of STK15/aurora2 kinase and involvement in colorectal cancers.
  - 41 Giet R, Uzbekov R, Cubizolles F, Le Guellec K, Prigent C: The xenopus laevis aurora related protein kinase pEq2 associates with and phosphorylates the Kinesin related protein X1Eq5. *J Biol Chem* 1999, 274:15005-15013.
  - 42 Zimmerman W, Sparks C, Duxsey S: Amorphous no longer: the centrosome comes into focus. *Curr Opin Cell Biol* 1998, 11:122-128.
  - 43 Zou H, McGarry TJ, Bernal T, Kirschner MW: Identification of a vertebrate sister chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* 1999, 285:418-421.  
Demonstrates transforming and tumorigenic function of a gene inhibiting sister chromatid separation.
  - 44 Hardwick KG: The spindle checkpoint. *Trends Genet* 1998, 14:1-4.
  - 45 Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JKV, et al.: Mutations of mitotic checkpoint genes in human cancers. *Nature* 1998, 392:300-303.  
Describes mitotic checkpoint gene mutations in human colorectal cancers showing chromosome instability.
  - 46 Jin DY, Spencer F, Jeang KT: Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell* 1998, 33:81-91.
  - 47 Xu X, Weaver Z, Linke SP, Li C, Gotay J, Wang XW, et al.: Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform deficient cells. *Mol Cell* 1999, 3:389-395.

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER: \_\_\_\_\_**

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**